

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 50.2179 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221b-10
Perfect score: 1340
Sequence: 1 GLKGRGDSGSPATWTTRGF.....KAGELEKIISRQVCMKKRH 244
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 1340 | 100.0 | 1670 | 1 CGH03B | collagen alpha 3(I) |
| 2 | 1218.5 | 90.9 | 246 | 2 I48302 | collagen alpha 3(I) |
| 3 | 1210.5 | 90.3 | 471 | 2 A39024 | collagen alpha 3(I) |
| 4 | 960 | 71.6 | 1669 | 1 CGH04B | collagen alpha 1(I) |
| 5 | 952.5 | 71.1 | 253 | 2 I48304 | collagen alpha 5(I) |
| 6 | 951 | 71.0 | 1669 | 1 CGM84B | collagen alpha 1(I) |
| 7 | 943.5 | 70.4 | 1691 | 1 S22917 | collagen alpha 1(I) |
| 8 | 943 | 70.4 | 258 | 2 B61228 | collagen alpha 1(I) |
| 9 | 922.5 | 68.8 | 754 | 2 A55267 | collagen alpha 5(I) |
| 10 | 872 | 65.1 | 220 | 2 B49736 | collagen alpha 5(I) |
| 11 | 852.5 | 63.6 | 1744 | 2 S40991 | collagen alpha 1(I) |
| 12 | 834 | 62.2 | 151 | 2 S49488 | collagen alpha 3(I) |
| 13 | 832 | 62.1 | 1782 | 2 A45407 | collagen alpha 3(I) |
| 14 | 796.5 | 59.4 | 1783 | 2 S16366 | collagen alpha 1(I) |
| 15 | 783.5 | 58.5 | 1758 | 2 T29350 | collagen alpha 5(I) |
| 16 | 783.5 | 58.5 | 1759 | 2 T29351 | collagen alpha 3(I) |
| 17 | 782 | 58.4 | 261 | 2 A34476 | collagen alpha 2(I) |
| 18 | 767.5 | 57.3 | 1747 | 2 A54121 | collagen alpha 4(I) |
| 19 | 760.5 | 56.8 | 1707 | 2 A33526 | collagen alpha 2(I) |
| 20 | 755.5 | 56.4 | 1712 | 1 CGH02B | collagen alpha 2(I) |
| 21 | 739 | 55.1 | 1691 | 1 CGH06B | collagen alpha 6(I) |
| 22 | 734.5 | 54.8 | 775 | 2 A61228 | collagen alpha 2(I) |
| 23 | 709 | 52.9 | 453 | 2 S18804 | collagen alpha 4(I) |
| 24 | 705 | 52.6 | 312 | 2 I48303 | collagen alpha 4(I) |
| 25 | 702 | 52.4 | 623 | 2 I48304 | collagen alpha 4(I) |
| 26 | 696 | 51.9 | 1690 | 1 CGH01B | collagen alpha 4(I) |
| 27 | 694.5 | 51.8 | 1775 | 2 A31893 | collagen alpha 1(I) |
| 28 | 646.5 | 48.2 | 1761 | 2 T13390 | collagen type IV a |
| 29 | 334 | 24.9 | 81 | 2 A49736 | collagen alpha 3(I) |

ALIGNMENTS

RESULT 1

CGH03B

collagen alpha 3(IV) chain precursor, long splice form - human
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 22-Jun-1999
C:Accession: A54763; A43928; A4043; A45971; A39786
R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Readers, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
A:Reference number: A54763; MUID:94364994; PMID:8083201
A:Accession: A54763
A:Molecule type: mRNA
A:Residues: 1-1670 <MAR>
A:Cross-references: GB:X80031; NID:G577563; PID:G577564
A:Experimental source: kidney
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the
A:Reference number: A43928; MUID:92147878; PMID:1737849
A:Accession: A43928
A:Molecule type: mRNA
A:Residues: 1331-1524, 1526-1670 <TUR>
A:Cross-references: GB:M81379
A:Experimental source: kidney
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
A:Reference number: A44043; MUID:93015826; PMID:1400291
A:Accession: A44043
A:Molecule type: DNA; mRNA
A:Residues: 1386-1670 <OUI>
A:Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896
A:Note: sequence extracted from NCEI backbone (NCBIP:115597)
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A:Reference number: A44738; MUID:94274734; PMID:8006044
A:Contents: annotation; erratum; correction to intronic sequence in A44043
R:Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A:Reference number: A45971; MUID:93280184; PMID:8505332
A:Accession: A45971
A:Molecule type: mRNA
A:Status: nucleic acid sequence not shown
A:Residues: 1427-1444 <BER>
A:Note: sequence extracted from NCEI backbone (NCBIP:133363); sequence incorrectly identified
R:Morrison, K.E.; Maruyama, M.; Yang-Feng, T.L.; Readers, S.T.
Am. J. Hum. Genet. 49, 545-554, 1991
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
A:Reference number: A39786; MUID:91353570; PMID:1882840

A:Accession: A39786
A:Molecule type: mRNA
A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
A:Cross-references: GB:S55790; NID:9234418; PID:NAB19637.1; PID:9234419
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C) and subsequently O-glycosylated.
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-2q37
A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3(IV) monomer trimers amino-terminal domains (with disulfide and desmosine cross-links), dimeric associations in the interrupted helical domain (with disulfide and desmosine cross-links).
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracellular matrix; signal sequence #status predicted <SIG>
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
F:39-42/Domain: amino-terminal nonhelical, NH1 <NH1>
F:43-1438/Region: interrupted helical
F:791-793/Region: cell attachment (R-G-D) motif
F:796-998/Region: cell attachment (R-G-D) motif
F:1154-1156/Region: cell attachment (R-G-D) motif
F:1306-1308/Region: cell attachment (R-G-D) motif
F:1345-1347/Region: cell attachment (R-G-D) motif
F:1432-1434/Region: cell attachment (R-G-D) motif
F:1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NCI>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:161-33, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1387/Disulfide bonds: interchain #status predicted
F:253/Binding site: carbonyl-Asn (covalent) #status predicted
F:1460-1548, 1493-1551/disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F:1505-1511, 1616-1622/disulfide bonds: #status predicted
F:1570-1662, 1604-1665/disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
Query Match 100.0%; Score 1340; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 4.8e-113;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFVQGNQRAHQD 60
DB 1427 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFVQGNQRAHQD 1486
QY 61 LGTLGSLQRFVTRHSQTTAIPSCPEGTVPLYSFVQGNQRAHQD 120
DB 1487 LGTLGSLQRFVTRHSQTTAIPSCPEGTVPLYSFVQGNQRAHQD 1546
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWTSLWKGFNFMTSAGSEGTQALASPGSCLE 180
DB 1547 RCTVCEGPAIAIAVHSQTTDIPPCPHGWTSLWKGFNFMTSAGSEGTQALASPGSCLE 1606
QY 181 EFRASPLFCHGRCNTYNSYSFVWLASLNFPMFRKPIPTSTVKGLEKIISRCQVC 240
DB 1607 EFRASPLFCHGRCNTYNSYSFVWLASLNFPMFRKPIPTSTVKGLEKIISRCQVC 1666
QY 241 MKRH 244
DB 1667 MKRH 1670
RESULT 2
collagen alpha 3(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
C:Accession: 148302; S47278
R:Winer, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-894, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A:Reference number: A54979; MUID:95050957; PMID:7962065
A:Accession: I48302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <RES>
A:Cross-references: EMBL:Z35166; NID:G535197; PID:G535198
C:Superfamily: collagen alpha 1(IV) chain
Query Match 90.9%; Score 1218.5; DB 2; Length 246;
Best Local Similarity 90.2%; Pred. No. 6.1e-103;
Matches 221; Conservative 12; Mismatches 11; Indels 1; Gaps 1;
QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFVQGNQRAHQD 59
DB 2 GLKGNFCDRTATGTRMGFIITRHSQTTAIPSCPEGTVPLYSFVQGNQRAHQD 61
QY 60 DLGTLGSLQRFVTRHSQTTDIPPCPHGWTSLWKGFNFMTSAGSEGTQALASPGSCLE 119
DB 62 DLGTLGSLQRFVTRHSQTTDIPPCPHGWTSLWKGFNFMTSAGSEGTQALASPGSCLE 121
QY 120 RCTVCEGPAIAIAVHSQTTDIPPCPHGWTSLWKGFNFMTSAGSEGTQALASPGSCLE 179
DB 122 RCTVCEGPAIAIAVHSQTTDIPPCPHGWTSLWKGFNFMTSAGSEGTQALASPGSCLE 181
QY 180 EFRASPLFCHGRCNTYNSYSFVWLASLNFPMFRKPIPTSTVKGLEKIISRCQVC 239
DB 182 EFRASPLFCHGRCNTYNSYSFVWLASLNFPMFRKPIPTSTVKGLEKIISRCQVC 241
QY 240 MKRH 244
DB 242 MKRH 246
RESULT 3
A39024
collagen alpha 3(IV) chain - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
R:Morrison, K.E.; Germino, G.G.; Reiders, S.T.
J. Biol. Chem. 266, 34-39, 1991
A:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the A:Reference number: A39024; MUID:91093146; PMID:1985905
A:Accession: A39024
A:Molecule type: mRNA
A:Residues: 1-471 <MOR>
A:Cross-references: EMBL:M63139; NID:G162886; PID:AAA62708.1; PID:G162887
R:Burkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J. Biol. Chem. 262, 7874-7877, 1987
A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane A:Reference number: S18432; MUID:97222419; PMID:2438283
A:Accession: S20672
A:Molecule type: protein
A:Residues: 227-228, 'X', 230-244 <BUT>
R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
J. Biol. Chem. 263, 13374-13380, 1988
A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen A:Reference number: S17802; MUID:88330844; PMID:3417561
A:Accession: S17802
A:Molecule type: protein
A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J. Biol. Chem. 265, 5466-5469, 1990
A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type A:Reference number: A35167; MUID:90202779; PMID:2318822
A:Accession: A35167
A:Molecule type: protein
A:Residues: 236-258 <GUN>
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Mc
J. Biol. Chem. 266, 15318-15324, 1991
A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol A:Reference number: A39419; MUID:91332055; PMID:1869555
A:Accession: C39419

A:Molecule type: protein
A:Residues: 236-255 <G02>
A:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F:1-238/Domain: collagenous (fragment) #status predicted <COL>
F:239-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC11>
F:239-353/Domain: repeat NCI #status predicted <NC11>
F:354-471/Domain: repeat NCI #status predicted <NC12>
F:232,238/Modified site: hydroxyproline (Pro) #status experimental
F:306-312,417-423/disulfide bonds: #status predicted

Query Match 90.3%; Score 1210.5; DB 2; Length 471;
Best Local Similarity 90.6%; Pred. No. 6.5e-102;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRKSGSPATWT-RGFVTRHSQTALPSCPEGVPLYSFSLFVQGNORAGQ 59
DB 227 GLKGRKSGSPATWT-RGFVTRHSQTALPSCPEGVPLYSFSLFVQGNORAGQ 286
QY 60 DLGTLGSLQRTTTPFLFCNVNDYCNFASNDYSYWLSTPALMPNMVAPITGRALEPYI 119
DB 287 DLGTLGSLQRTTTPFLFCNVNDYCNFASNDYSYWLSTPALMPNMVAPITGRALEPYI 346
QY 120 SRCTWCEGPAIAIAVHSQTTDTPCPHGWISLWKGFSPMTFSAGSEGTGQALASPGSCL 179
DB 347 SRCTWCEGPAIAIAVHSQTTDTPCPHGWISLWKGFSPMTFSAGSEGTGQALASPGSCL 406
QY 180 EEFRAFPFTECHGRGFCNYNSYSFPLASLNPFRMFKPIPTSTVKAGELEKIISRCQVC 239
DB 407 EEFRAFPFTECHGRGFCNYNSYSFPLASLNPFRMFKPIPTSTVKAGELEKIISRCQVC 466
QY 240 MKKR 243
DB 467 MKKR 470

RESULT 4
CGHU4B
collagen alpha 1(IV) chain precursor - human
N:Alternate names: procollagen alpha 1(IV) chain
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999
C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
R:Soiainen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
A:Reference number: S16876; MUID:89340433; PMID:2701944
A:Accession: S16876
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1669 <SO11>
A:Cross-references: EMBL:J04217; GB:J05039; NID:G180759; PIDN:AAA53097.1; PID:G553233
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
R:Soiainen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A:Reference number: A92690; MUID:89034231; PMID:3182844
A:Accession: A32117
A:Molecule type: DNA
A:Residues: 1-28 <SO12>
A:Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233
R:Peeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c
A:Reference number: S02738; MUID:89030632; PMID:2846280
A:Accession: S02738
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-6, 'L', 8-28 <POE>
A:Cross-references: EMBL:X12784; NID:G30072
R:Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem

A:Reference number: S00048; MUID:88029471; PMID:3311751
A:Accession: S00048
A:Molecule type: mRNA
A:Residues: 1-318, 'A', 320-944 <BRA1>
A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067
A:Accession: S25826
A:Molecule type: protein
A:Residues: 271-318, 'A', 320-554 <BRA2>
R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7;
A:Reference number: A23115; MUID:86004708; PMID:4043082
A:Accession: A23115
A:Molecule type: protein
A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>
A:Experimental source: placenta
A:Note: the amino end of the mature form is blocked
R:Soiainen, R.; Raka-Riska, T.; Prockop, D.J.; Tryggvason, K.
FEBS Lett. 225, 188-194, 1987
A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane (;
A:Reference number: S00207; MUID:88083584; PMID:3691802
A:Accession: S00207
A:Molecule type: mRNA
A:Residues: 244-530 <SO13>
A:Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549
R:Edle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collage;
A:Reference number: S39614; MUID:94038963; PMID:8223488
A:Accession: S39614
A:Molecule type: protein
A:Residues: 371-554 <EBL>
R:Babel, W.; Glanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid ;
A:Reference number: A02863; MUID:85003629; PMID:6434307
A:Accession: A02863
A:Molecule type: protein
A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 99;
A:Experimental source: placenta
R:Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A:Title: Pepsin fragments of human placental basement-membrane collagens showing inter;
A:Reference number: S16908; MUID:82005835; PMID:6792033
A:Accession: A58517
A:Molecule type: protein
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'XX', 1408-1409, 'X', 1411-;
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane
A:Reference number: S16910; MUID:84033346; PMID:6416291
A:Accession: S16910
A:Molecule type: protein
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'Y', 946, 'X', 94;
A:Experimental source: placenta
R:Philiangemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;
J. Biol. Chem. 260, 7681-7687, 1985
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
A:Reference number: S01466; MUID:85207819; PMID:2581969
A:Accession: S01466
A:Molecule type: mRNA
A:Residues: 1256-1669 <PIH>
A:Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424
R:Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chain;
A:Reference number: S16879; MUID:85216555; PMID:2582422
A:Accession: S16879
A:Molecule type: mRNA
A:Residues: 1259-1669 <BRI>
A:Cross-references: EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PID:G180818
R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss
Eur. J. Biochem. 147, 217-224, 1985

```

Db      1547 RCACVEAPAMVAVHSQTIIQPPCCSGWSLMIIGYSFVYHMTSAGAGSGQALASPGSCLE 1500
QY      181 EFRASPFLECHGRGTCNYYSNSYSFWLASLNPMPERFKPIPTSVTKAGELEKIISRQVCM 240
Db      1607 EFRSAPFIECHGRGTCNYYANAYSFWLAIERSEMEFKPTPTSLKAGELRTHYSRCQVCM 1666
QY      241 KK 242
Db      1667 RR 1668

RESULT 5
I4304
Collagen alpha 5(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999
C:Accession: I43034; S47280
R:Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal lamin
A:Reference number: A54979; MUID:95050957; PMID:7962065
A:Accession: I48304
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <RES>
A:Cross-references: EMBL:235168; NID:G535201; PIDN:CAA84531.1; PID:G535202
C:Superfamily: collagen alpha 1(IV) chain

Query Match      71.1%; Score 952.5; DB 2; Length 253;
Best Local Similarity 68.7%; Pred. No. 7.9e-79;
Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY      1 GLKKGKGDGSGPATWT-TRGFVTRHSQTATPSCPEGTVPVLYSGFSFLVQGNQRAHQ 59
Db      10 GPDGLGPPGPGTTSVAHGFLITRSQTEAPQCPRGVVHYTEGFSLLFVQGNKRAHQ 69
QY      60 DLGTGLSCLORFMTTFLEFCNVNDVCNFASRNDYSYWLSTPALPMNMAPITGRALPEYI 119
Db      70 DLGTAGSCLRRFTSMFPMFCNNVCNFASRNDYSYWLSTPEPMNMEPLKGSIQPFI 129
QY      120 SRCTVCEGPAIAVAHSQTTDPPCPHGWSLWKGFESFMFTSAGSGTGOALASPGSCL 179
Db      130 SRCACVEAPAVJAVHSQTIQPHCPQGWSLMIIGYSFVYHMTSAGAGSGQALASPGSCL 189
QY      180 EFRSAPFIECHGRGTCNYYSNSYSFWLASLNPMPERFKPIPTSVTKAGELEKIISRQVC 239
Db      190 EFRSAPFIECHGRGTCNYYANSYSFWLATVDMDFNKPEQSETLAKAGLRTRISRCQVC 249
QY      240 MKK 242
Db      250 MKR 252

RESULT 6
CGMS48
Collagen alpha 1(IV) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
C:Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A32003; A3
R:Muthukumar, G.; Blumberg, B.; Kurkinen, M.
J. Biol. Chem. 264, 6310-6317, 1989
A:Title: The complete primary structure for the alpha-1-chain of mouse collagen
A:Reference number: A33525; MUID:89197932; PMID:2703490
A:Accession: A33525
A:Molecule type: mRNA
A:Residues: 1-1669 <MUT>
A:Cross-references: EMBL:J04694; NID:G556296; PIDN:AAA50292.1; PID:G556297
R:Wood, L.; Thieriault, N.; Vogeli, G.
FEBS Lett. 227, 5-8, 1988
A:Title: cDNA clones completing the nucleotide and derived amino acid sequence
A:Reference number: S01454; MUID:8811221; PMID:3338568
A:Accession: S01454
A:Molecule type: mRNA

```

A;Residues: 940-946,'G',948-949,'C',951-955,'G',957-964,'X',966-991,'X',993-1003,'X',1003-1011,'X',1063-1065,'X',1067-1080,'X',1085-1106,'X',1108-1115,'DE',1118-1111

Eur. J. Biochem. 123, 505-512, 1982

A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial am

A;Reference number: A25991; MUID:82186723; PMID:6804236

A;Accession: A25991

A;Molecule type: protein

A;Residues: 940-946,'G',948-949,'X',951-955,'X',957-964,'X',966-991,'X',993-1003,'X',1003-1011,'X',1063-1065,'X',1067-1080,'X',1085-1106,'X',1108-1115,'DE',1118-1111

A;Accession: B25991

A;Molecule type: protein

A;Residues: 1173-1181,'X',1183-1184,'X',1186-1187,'X',1189-1205,'Q',1207,'XE',1210-1213,'3','Sp',1266,'IT',1269,'SK',1272,'DW',1275,'L',1277-1282,'1316-1318,'X',1320-1327,'X',1313,'Weber, S., Engel, J., Wiedemann, H.; Glanville, R.W.; Timpl, R.

Eur. J. Biochem. 139, 401-410, 1984

A;Title: Subunit structure and assembly of the globular domain of basement-membrane co

A;Reference number: S17801; MUID:84132058; PMID:6698021

A;Accession: S17801

A;Molecule type: protein

A;Residues: 1435-1443 <WEB>

C;Genetics:

A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3

A;Note: the list of introns may be incomplete

C;Superfamily: collagen alpha 1(IV) chain

C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular m

F;1-27/Domain: signal sequence #status predicted <SIG>

F;28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>

F;28-162/Domain: 78 <7SD>

F;163-1440/Domain: collagenous, triple helix <COL>

F;597-599/Region: cell attachment (R-G-D) motif

F;781-783/Region: cell attachment (R-G-D) motif

F;917-919/Region: cell attachment (R-G-D) motif

F;968-970/Region: cell attachment (R-G-D) motif

F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

F;1441-1659/Region: duplication

F;1553-1559/Region: duplication

F;31,36,39,41,43,46,47,470/Disulfide bonds: interchain #status predicted

F;126/Binding site: carbohydurate (Asn) (covalent) #status predicted

F;971,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,1090

92,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxyp

F;1214,1424/Modified site: 4-hydroxyproline (pro) #status experimental

F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental

F;1305-1511,1616-1622/Disulfide bonds: #status predicted

Query Match 71.0%; Score 951; DB 1; Length 1669;

Best Local Similarity 68.6%; Pred. No. 8.3e-78;

Matches 166; Conservative 34; Mismatches 40; Indels 2; Gaps 1;

Qy 1 GLKRGDGSQSPATWTTTRGGVFTBHSOTTATPSCPEGTVPLYSQSFELFVQGNRAHQD 60

Db 1429 GLPGSMGPPGTPGS--VDHGLFVTRHSQTTDPLCPGPKILYHGYSLLYVQGNRAHQD 1486

Qy 61 LGTLGSLQRFPTTFMPFFLCNVNDYCNFASRNDYSYWLSTPALPMNMAPITGRALEPYIS 120

Db 1487 LGTAGSCLRFSTWPFELFCNINVCNFRNDYSYWLSTPEPMPMAPISGDNIRPFI 1546

Qy 121 RCTVCEGPAIAVHSQTTDIPPCPHGWSLWKGSFIMFTSAGEGQALASPGSCLE 180

Db 1547 RCACVEAPAMVMVHSQTIQIQCPCNGWSSLIWISYFVMHTSAGAGSGQALASPGSCLE 1606

Qy 181 EFRASPLECHGRGTCNYSNSYFVLASLNPESMFRKPIPTSVKAGSELEKIISRCQVCM 240

Db 1607 EFRSAPLECHGRGTCNYSNAYSFWLATIERSEMFKFTSTLKAGELRTHVSRCQVCM 1666

Qy 241 KK 242

Db 1667 RR 1668

RESULT 7

S22917

collagen alpha 5(IV) chain precursor, renal splice form - human

N;Alternate names: procollagen alpha 5(IV) chain

N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
 C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text_change 21-Jul-2000
 A:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A39
 J: Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
 J. Biol. Chem. 267, 12475-12481, 1992
 A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi
 n Alport syndrome patient.
 A:Reference number: S22917; MUID:93216923; PMID:1352287
 A:Accession: S22917
 A:Molecule type: mRNA
 A:Residues: 1-967 <ZHO>
 A:Cross-references: GB:M90464; NID:q180826; PIDN:AAA52046.1; PID:g553234
 R: Zhou, J.; Leinonen, A.; Tryggvason, K.
 J. Biol. Chem. 269, 6608-6614, 1994
 A:Title: Structure of the human type IV collagen COL4A5 gene.
 A:Accession: A54365
 A:Reference number: A54365; MUID:94165049; PMID:8120014
 A:Molecule type: DNA
 A:Residues: 1-922 <ZH2>
 A:Cross-references: GB:U04470; NID:q463378; GB:U04520; NID:q463428; PIDN:AAC27816.1; PID
 R: Zhou, J.; Mochizuki, T.; Sneets, H.; Antignac, C.; Laurila, P.; de Paeppe, A.; Tryggvason
 Science 261, 1167-1169, 1993
 A:Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited sc
 A:Reference number: A57079; MUID:93361972; PMID:8356449
 A:Accession: A57079
 A:Molecule type: DNA
 A:Residues: 1-27 <ZH4>
 A:Cross-references: GB:Z37153; NID:q587203; PIDN:CAA85512.1; PID:g587204
 R: Pihlajaniemi, T.; Pihlajaniemi, E.R.; Myers, J.C.
 J. Biol. Chem. 265, 13758-13766, 1990
 A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin
 A:Reference number: A37122; MUID:90337990; PMID:2380186
 A:Accession: A37122
 A:Molecule type: mRNA
 A:Residues: 84-439, 'GS', 442-624, 'LALO', 629-665, 'ER', 669-887, 'R', 889-1264, 1271-1691 <PIH>
 A:Cross-references: GB:J05558; EMBL:M58526; NID:q1314209
 A:Note: submitted to the EMBL Data Library, February 1991
 A:Note: the authors translated the codon GCC for residue 115 as Val
 R: Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
 Hum. Mol. Genet. 1, 127-129, 1992
 A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
 A:Reference number: I54317; MUID:93244772; PMID:1363780
 A:Accession: I54317
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 313-324, 'E', 326-330 <REN>
 A:Cross-references: GB:S59334; NID:q299946; PIDN:AAD13909.1; PID:g4261609
 R: Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Tryggvason, K.
 Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
 A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidne
 A:Reference number: A34850; MUID:90160375; PMID:1689491
 A:Accession: A34850
 A:Molecule type: mRNA
 A:Residues: 914-1264, 1271-1691 <HOS>
 A:Cross-references: EMBL:M31115; NID:q180824; PIDN:AAA52045.1; PID:g180825
 R: Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.
 Genomics 9, 1-9, 1991
 A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene tha
 A:Reference number: A37969; MUID:91169491; PMID:2004755
 A:Accession: S18850
 A:Molecule type: DNA
 A:Residues: 924-1264, 1271-1691 <ZH3>
 A:CROSS-references: EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB
 8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:q177922; PIDN:AAA51558.1; PID
 R: Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M
 Kidney Int. 44, 1316-1321, 1993
 A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
 A:Reference number: I56971; MUID:94133540; PMID:8301933
 A:Accession: I56971
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1258-1276 <GUO1>

A:Cross-references: GB:S63168; NID:g545095; PIDN:AAC60612.1; PID:g545096
 A:Note: kidney splice form
 A:Accession: I76598
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1284-1291, 'TFLGYLACLIV' <GUO2>
 A:Cross-references: GB:S69169; NID:g545097; PIDN:AAC60613.1; PID:g545098
 A:Note: frameshift mutation in patient with Alport syndrome
 R: Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; S
 Am. J. Hum. Genet. 46, 1024-1033, 1990
 A:Title: Molecular cloning of alphas(IV) collagen and assignment of the gene to the reg
 A:Reference number: A35335; MUID:90252791; PMID:2339699
 A:Accession: A35335
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1448-1477 <MYE>
 R: Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Yc
 Kidney Int. 46, 1307-1314, 1994
 A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primor
 A:Reference number: I56975; MUID:95156893; PMID:7853788
 A:Accession: I56975
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1595-1602 <NAK>
 A:Cross-references: GB:S75903; NID:q913882; PIDN:AAB33374.1; PID:g913883
 A:Note: premature termination mutation from a patient with Alport syndrome; one other n
 R: Lemmink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;
 Genomics 17, 485-489, 1993
 A:Title: Identification of four novel mutations in the COL4A5 gene of patients with Alf
 A:Reference number: I54188; MUID:94010948; PMID:8406498
 A:Accession: I54188
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1604-1607, 'VHDAYKC' <LEM>
 A:Cross-references: GB:S65767; NID:q425563; PIDN:AAD13967.1; PID:g4261667
 A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations e
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: COL4A5; ATS
 A:Cross-references: GDB:120596; OMIM:303630
 A:Map position: Xq22-Xq22
 A:Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229
 /3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1;
 A:Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands wit
 C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
 er associations in the interrupted helical domain (with disulfide and desmosine cross-l
 C:Function:
 A:Description: minor structural component of extracellular basement membrane
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; gly
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAI
 F:27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status
 F:27-1462/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
 F:27-1462/Region: interrupted helical
 F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
 F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
 F:125/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
 F:1527-1533,1538-1644/Disulfide bonds: #status predicted
 F:1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted
 Query Match 70.4%; Score 943.5; DB 1; Length 1691;
 Best Local Similarity 67.9%; Pred. No. 4e-77;
 Matches 165; Conservative 35; Mismatches 42; Indels 1; Gaps 1;
 Qy 1 GLKKGCDGSPATWT-TRGPFVTRHSQTATPSCPEGTVPLYSGFSLFVGNORAHQ 59
 Db 1448 GPDGLQGPFGPTSSVAMGFLTRHSQRTDAPQCQFLQVYEGSLLYVGNKRAHQ 1507

QY 60 DLGTGLGSLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPNMVAPITGRALEPYI 119
DB 1508 DLGTAGSCLRRFSTMPFECNINNVCFASRNDYSYWLSTPEPMPMSQPLKGOSIQPFI 1567
QY 120 SRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCL 179
DB 1568 SRCVACEAPAVVIAVHSQTIQIHPCPQGWDSLWATGYSPMHTSAGSEGTGQALASPGSCL 1627
QY 180 EFRASPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIPTSTVKAGELEKIISRCQVC 239
DB 1628 EFRSAPFIECHGRGTCNYNSYSYFWLATVDVDSMFSPKQSETLKAGDLRTRISRCQVC 1687
QY 240 MKK 242
DB 1688 MKR 1690

RESULT 8

B61228
collagen alpha 1(IV) chain - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
C:Accession: B61228
R.Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991
A:Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium
A:Reference number: A61228; MUID:92010685; PMID:1717398
A:Accession: B61228
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-258 <YAM>
C:Superfamily: collagen alpha 1(IV) chain

Query Match 70.4%; Score 943; DB 2; Length 258;
Best Local Similarity 67.8%; Pred. No. 5.8e-78;
Matches 164; Conservative 34; Mismatches 42; Indels 2; Gaps 1;
QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQGD 60
DB 18 GLPGSGMPGPTFS--VDHGFLVTRHSQTTDHPQCPGPKILYHGYSLLVVGNERAHQGD 75
QY 61 LGTLGSLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPNMVAPITGRALEPYIS 120
DB 76 LGTAGSCLRKFTMPFLFCNINNVCFASRNDYSYWLSTPEPMPMSMAPMWDNIRPFI 135
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCL 180
DB 136 RCVACEAPAVVIAVHSQTIQIHPCPNGWSSLWATGYSPMHTSAGSEGTGQALASPGSCL 195
QY 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIPTSTVKAGELEKIISRCQVC 240
DB 196 EFRSAPFIECHGRGTCNYNSYSYFWLATIERSEMFKPTPTSLKAGELHTHVSRCQVC 255
QY 241 KK 242
DB 256 RR 257

RESULT 9

A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267
R.Zheng, K.; Thorne, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked type IV.
A:Reference number: A55267; MUID:94224868; PMID:8171024
A:Accession: A55267
A>Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>
A:Cross-references: GB:U07888; NID:G469547; PIDN:AA860258.1; PID:G469548
C:Superfamily: collagen alpha 1(IV) chain

Query Match 68.8%; Score 922.5; DB 2; Length 754;
Best Local Similarity 68.4%; Pred. No. 1.3e-75;
Matches 162; Conservative 32; Mismatches 42; Indels 1; Gaps 1;
QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQGD 59
DB 518 GPDGMGPPGPTGTSIAHGFLTRHSQTTDAPQPHGTIVQIYEGFSLLYVQGNORAHQGD 577
QY 60 DLGTGLGSLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPNMVAPITGRALEPYI 119
DB 578 DLGTAGSCLRRFSTMPFECNINNVCFASRNDYSYWLSTPEPMPMSMEPLKGQSTQTFPI 637
QY 120 SRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCL 179
DB 638 SRCVACEAPAVVIAVHSQTIQIHPCHPGWDSLWATGYSPMHTSAGSEGTGQALASPGSCL 697
QY 180 EFRASPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIPTSTVKAGELEKIISRC 236
DB 698 EFRSAPFIECHGRGTCNYNSYSYFWLATVDVDSMFSPKQSETLKAGDLRTRISRC 754

RESULT 10

B49736
collagen alpha 3(IV) chain, medium splice form - human (fragment)
N:Contains: collagen alpha 3(IV) chain, splice form GP-V
C:Species: Homo sapiens (man)
C>Date: 03-May-1994 #sequence_revision 12-Nov-1999 #text_change 17-Mar-2000
C:Accession: B49736; S69111
R.Feng, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A:Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.
A:Reference number: A49736; MUID:94124597; PMID:8294492
A:Accession: B49736
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 169-220 <PEN1>
A:Accession: B49736
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 22-220 <PEN2>
A:Cross-references: GB:U02519; NID:G409106; PIDN:AA18942.1; PID:G409107
A:Note: This is the conceptual translation of the nucleic acid submitted to GenBank
R:Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; W.
Eur. J. Biochem. 229, 754-760, 1995
A:Title: Characterization and expression of multiple alternatively spliced transcripts
uot antigen and one of its alternative forms
A:Reference number: S69111; MUID:95278230; PMID:7758473
A:Accession: S69111
A:Molecule type: mRNA
A:Residues: 1-45,169-204,'L',206-220 <PEN>
C:Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-2q37
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extract
F:1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status prec
F:1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status
F:22-220/Dominant: carboxyl-terminal nonhelical, NC1 <NC1>
F:34-134/Dominant: collagen IV carboxyl-terminal repeat <CTL>

Query Match 65.1%; Score 872; DB 2; Length 220;
Best Local Similarity 99.4%; Pred. No. 1.3e-71;
Matches 158; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQGD 60
DB 10 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQGD 69

QY 61 LGTGLSCLOQTTPFPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALPEYIS 120
DB 70 LGTGLSCLOQTTPFPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALPEYIS 129
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSEIM 159
DB 130 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSEIM 168
RESULT 11
S40991
collagen alpha 1(IV) chain precursor - Caenorhabditis elegans
N:Alternate names: protein K04H4.1
C:Species: Caenorhabditis elegans
C:Date: 03-May-1994 #sequence_revision 02-Aug-1994 #text_change 13-Aug-1999
R:Ainscough, R.
C:Accession: S40991; S44442; S13651; B34476
submitted to the EMBL Data Library, October 1993
A:Reference number: S40991
A:Molecule type: DNA
A:Residues: 1-1744 <AIN>
A:Cross-references: EMBL:X27078; NID:G414627; PID:G414628
R:Kramer, J.M.
submitted to the EMBL Data Library, December 1990
A:Reference number: S44442
A:Accession: S44442
A:Molecule type: DNA
A:Residues: 1-129 'GFGQMPGLAGPQSGQNGNPGRLGSLGPPGEGVNSQGRKGKVGESGVPGLP', 209-281, 'PV
15, 'D', 817-1260, 'P', 1262-1707, 'P', 1709-1744 <KRA>
A:Cross-references: EMBL:X56979; NID:G6675; PIDN:CAA40299.1; PID:G6676
R:Guo, X.; Johnson, J.J.; Kramer, J.M.
Nature 349, 707-709, 1991
A:Title: Embryonic lethality caused by mutations in basement membrane collagen of C. ele
A:Reference number: S13651; MUID:91141582; PMID:1996137
A:Accession: S13651
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-129 'GFGQMPGLAGPQSGQNGNPGRLGSLGPPGEGVNSQGRKGKVGESGVPGLP', 209-281, 'PV
15, 'D', 817-1260, 'P', 1262-1515 <GU1>
A:Cross-references: EMBL:X56979
R:Guo, X.; Kramer, J.M.
J. Biol. Chem. 264, 17574-17582, 1989
A:Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1
A:Reference number: A34476; MUID:90008929; PMID:2793871
A:Accession: B34476
A:Molecule type: DNA
A:Residues: 1432-1499, 'Q', 1501-1707, 'P', 1709-1744 <GU2>
A:Cross-references: EMBL:X05067; NID:G156255; PIDN:AB59179.1; PID:G156256
C:Genetics:
A:Gene: clb-2, emb-9
A:Map position: 3
A:Interons: 23/2; 79/1; 152/2; 289/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F:43-1515/Domain: collagen; triple helix #status predicted <COL>
F:93-95/Region: cell attachment (R-G-D) motif
F:1053-1055/Region: cell attachment (R-G-D) motif
F:1396-1398/Region: cell attachment (R-G-D) motif
F:1516-1744/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>
F:1516-1627, 1628-1744/Region: duplication
F:1580-1586, 1691-1697/Disulfide bonds: #status predicted
Query Match 63.6%; Score 852.5; DB 2; Length 1744;
Best Local Similarity 60.7%; Pred. No. 7.2e-69;
Matches 148; Conservative 35; Mismatches 60; Indels 1; Gaps 1;
QY 1 GLKXKXGSGSPATWT-TRGFVTRHSQTTAIPSCPEGVTPLYSGFSLFVQGNQRAHQ 59
DB 1501 GLPPTGYGSGCGWAPSGGF-FAKHISQTTAVPQPPGASQLWEGYSLLYVQNGRASGQ 1560
QY 60 DLGTLGSCLOQTTPFPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALPEYI 119

DB 1561 DLGQFGSCLSKENTPFPCFNWNSVCHVSSRNDISFWLSTDEPMTPMNPTGTAIRPYI 1620
QY 120 SRTVCCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSEIMTSAGSEGTGQALASPGSCL 179
DB 1621 SRCVCEVPTQIIAIVHSQDTSVPQCPQGWGSMWTGYSFVMTAAAGETGQSLQSPGSCL 1680
QY 180 EEFRAFPFLECHGRGTCNVYNSYFWLASLNPFRMFKPIPSVTKAGELKIIISRCQVC 239
DB 1681 EEFRAVPFIECHGRGTCNYATNHGFWLSIVDQDKFKPKMSQTLKAGGLKDRVSRQVC 1740
QY 240 MKKR 243
DB 1741 LKNR 1744
RESULT 12
S49488
collagen alpha 3(IV) chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C:Accession: S49488
R:Oberbauer, I.
submitted to the EMBL Data Library, October 1994
A:Description: Cloning of the NCI domains to the minor collagen IV chains of mouse via
ells.
A:Reference number: S49487
A:Accession: S49488
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-161 <CBE>
A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G559916
C:Superfamily: collagen alpha 1(IV) chain
Query Match 62.2%; Score 834; DB 2; Length 161;
Best Local Similarity 92.5%; Pred. No. 2.6e-68;
Matches 149; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
QY 66 SCLQRTTTPFPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALPEYISRCTVC 125
DB 1 SCLQRTTTPFPFLFCNNVNCNFASRNDYSYWLSTPALMPMDNAPISGRALPEYISRCTVC 60
QY 126 EGPAIAIAVHSQTTDIPPCPHGWISLWKGFSEIMTSAGSEGTGQALASPGSCLBEFRAS 185
DB 61 EGPAIAIAVHSQTTAIPPCPDWVSLWKGFSEIMTSAGSEGTGQALASPGSCLBEFRAS 120
QY 186 PFLECHGRGTCNVYNSYFWLASLNPFRMFKPIPSVTYKA 226
DB 121 PFTECHGRGTCNVYNSYFWLASLNPFRMFKPIPSVTYKA 161
RESULT 13
A45407
collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A45407; A43903; A23940
R:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.
J. Biol. Chem. 268, 5249-5254, 1993
A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and an
A:Reference number: A45407; MUID:93186842; PMID:8444899
A:Accession: A45407
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1752 <EXP>
A:Note: sequence extracted from NCBI backbone (NCBI:P126841)
R:Wessel, G.M.; Etkin, M.; Benson, S.
Dev. Biol. 148, 261-272, 1991
A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously proc
A:Reference number: A43903; MUID:92038439; PMID:1936564
A:Accession: A43903
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 'P', 633-1537, 'G' <WES>
A:Cross-references: GB:S64572; NID:G238616; PIDN:AA820270.1; PID:G938617
A:Note: sequence extracted from NCBI Backbone (NCBIN:64572, NCBIP:64573)
R:Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
A:Title: Structure and developmentally regulated expression of a Strongylocentrotus purpuratus collagen alpha 2(IV) chain
A:Reference number: A23940; MUID:86205894; PMID:3455186
A:Accession: A23940
A:Molecule type: DNA
A:Residues: 742-812 <VEN>
A:Cross-references: ENBL:M13206
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:29-161/Domain: amino-terminal nonhelical, 7S <7SD>
F:162-1523/Region: interrupted helical
F:1524-1752/Domain: carboxyl-terminal nonhelical, NCI <NCI>
F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT>
F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT>
F:125/Modified site: allylsine (Lys) #status predicted

Query Match 62.1%; Score 832; DB 2; Length 1752;
Best Local Similarity 59.3%; Pred. No. 5.2e-67;
Matches 144; Conservative 37; Mismatches 60; Indels 2; Gaps 1;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60
Db 1512 GPGCTKGEAGIPG--SSSGFTTRHSQTTSIPQCPQGTAKMWHGYSLLFVQNERGHQD 1569

QY 61 LGTGLSCLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 120
Db 1570 LGKPGSCLURSTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 1629

QY 121 RCTVCEGPAIAVHSGTDTTPCPHGWSLWKGSFTMTFSAEGTGQALASPGSCLE 180
Db 1630 RCWCEAPAVLTVHSGTDTTPCPDRGVLWIGYSFWHTGPGEGSGQMLSSPGSCLE 1689

QY 181 EFRASPFLECHGRGTCNYSYNSYFWLASLNPFRMFKPIPTVTKAGELEKIIISRCQVC 240
Db 1690 DFRSPFTECHGRGTCNYSYNSYFWLASLNPFRMFKPIPTVTKAGELEKIIISRCQVC 1749

QY 241 KKR 243
Db 1750 RNQ 1752

RESULT 14
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C:Species: Ascaris suum (pig roundworm)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: S16366
R:Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the part
A:Reference number: S16366; MUID:913340768; PMID:1714907
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <JBI>
A:Cross-references: GB:M67507; NID:G159648; PIDN:AAA18014.1; PID:G159649
C:Genetics:
A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
A:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid
F:1-26/Domain: signal sequence #status predicted <Sig>
F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F:43-1529/Domain: collagenous #status predicted <COL>
F:197-199/Region: cell attachment (R-G-D) motif
F:1530-1763/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
F:1530-1638/Domain: repeat NCI #status predicted <NCI1>
F:1639-1763/Domain: repeat NCI #status predicted <NCI2>
F:31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 59.4%; Score 796.5; DB 2; Length 1763;
Best Local Similarity 57.9%; Pred. No. 8.7e-64;
Matches 140; Conservative 41; Mismatches 58; Indels 3; Gaps 2;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60
Db 1515 GLPGSGPPGPPGPKYKDGFLLVKHSQISEVPQCPGPMVKLWDGYSLLYIEGNEKSHNQD 1574

QY 61 LGTGLSCLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 120
Db 1575 LGHAGSCLSRFTSTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 1632

QY 121 RCTVCEGPAIAVHSGTDTTPCPHGWSLWKGSFTMTFSAEGTGQALASPGSCLE 180
Db 1633 RCACVEAPANVAVHSGTDTTPCPHGWSLWKGSFTMTFSAEGTGQALASPGSCLE 1692

QY 181 EFRASPFLECHGRGTCNYSYNSYFWLASLNPFRMFKPIPTVTKAGELEKIIISRCQVC 239
Db 1693 DFRATPFIECHGRGTCNYSYNSYFWLASLNPFRMFKPIPTVTKAGELEKIIISRCQVC 1752

QY 240 MK 241
Db 1753 IR 1754

RESULT 15
T29350
hypothetical protein F01G12.5a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29350
R:Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F01G12.
A:Reference number: Z20611
A:Accession: T29350
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1758 <WUX>
A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a
A:Experimental source: strain Bristol N2; clone F01G12
C:Genetics:
A:Gene: CESP:F01G12.5a
A:Map position: X
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3;
C:Superfamily: collagen alpha 1(IV) chain

Query Match 58.5%; Score 783.5; DB 2; Length 1758;
Best Local Similarity 55.3%; Pred. No. 1.3e-62;
Matches 142; Conservative 41; Mismatches 59; Indels 15; Gaps 4;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSLFVQ 51
Db 1504 GLDQPGGPGAGLPGAGAGPARDGFLVKHSQTTEVRCPEGQTKLWDGYSLLYIE 1563

QY 52 GNORAHQDGLTGLSCLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMNAPIT 111
Db 1564 GNEKSHNQDLGAGSCLQRTTTPPFLFCNVNDVCFASRNDYSYWLSTSEAIIP--MMFVN 1621

QY 112 GRALPYSRVCVCGGPAIAVHSGTDTTPCPHGWSLWKGSFTMTFSAEGTGQ 171
Db 1622 EREIPYSRVCVCGGPAIAVHSGTDTTPCPHGWSLWKGSFTMTFSAEGTGQ 1681

QY 172 LASPGSCLSEFPASPLECHGRGTCNYSYNSYFWLASLNPFRMFKPIPTVTKAGELE 230
Db 1682 LSSPGSCLSEFPATPFIECHGRGTCNYSYNSYFWLASLNPFRMFKPIPTVTKAGELE 1741

QY 231 KIISRCQVCMMK---RH 244
Db 1742 TRVSRQVCVKSTDGRH 1758

Search completed: April 5, 2004, 07:05:33
Job time : 51.2179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 30.7215 Seconds
(without alignments)
413.557 Million cell updates/sec

Title: US-10-032-221B-10
Perfect score: 1340
Sequence: 1 GLKRGDSGPATWTRGP.....KAGELEKIIISRCQVCMKKKH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|---------------|--------------------|
| 1 | 1340 | 100.0 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 2 | 1210.5 | 90.3 | 471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 3 | 960 | 71.6 | 1669 | 1 CA14 HUMAN | P02462 mus sapien |
| 4 | 951 | 71.0 | 1669 | 1 CA14 MOUSE | P02463 mus sapien |
| 5 | 943.5 | 70.4 | 1685 | 1 CA54 HUMAN | P29400 homo sapien |
| 6 | 922.5 | 68.8 | 754 | 1 CA54 CANFA | Q28247 canis famil |
| 7 | 845.5 | 63.1 | 1758 | 1 CA14 CAEEL | P17139 caenorhabdi |
| 8 | 795.5 | 59.4 | 1763 | 1 CA24 ASGSU | P27333 ascaris suu |
| 9 | 772.5 | 57.6 | 1758 | 1 CA24 CAEEL | P17140 caenorhabdi |
| 10 | 760.5 | 56.8 | 1707 | 1 CA24 MOUSE | P08122 mus musculu |
| 11 | 755.5 | 56.4 | 1712 | 1 CA24 HUMAN | P08572 homo sapien |
| 12 | 739 | 55.1 | 1691 | 1 CA64 HUMAN | Q14031 homo sapien |
| 13 | 709 | 52.9 | 453 | 1 CA44 BOVIN | Q29442 bos taurus |
| 14 | 702 | 52.4 | 623 | 1 CA44 RABIT | P55787 cryptolagus |
| 15 | 696 | 51.9 | 1690 | 1 CA44 HUMAN | P53420 homo sapien |
| 16 | 694.5 | 51.8 | 1775 | 1 CA14 DROME | P08120 drosophila |
| 17 | 98.5 | 7.2 | 4391 | 1 COX1 SCHPO | P07657 schizosacch |
| 18 | 96.5 | 7.4 | 4391 | 1 PGEM HUMAN | P38160 homo sapien |
| 19 | 89.5 | 6.7 | 3707 | 1 PGEM MOUSE | Q05793 mus musculu |
| 20 | 86.5 | 6.5 | 755 | 1 MTS1 HUMAN | O43312 homo sapien |
| 21 | 81.5 | 6.1 | 539 | 1 IL2B MOUSE | P16297 mus musculu |
| 22 | 80 | 6.0 | 599 | 1 G363 LEICH | P15706 leishmania |
| 23 | 79.5 | 5.9 | 1597 | 1 M3K4 MOUSE | Q08648 mus musculu |
| 24 | 79 | 5.9 | 837 | 1 GCSR MOUSE | P40223 mus musculu |
| 25 | 79 | 5.9 | 1877 | 1 PK5 MOUSE | Q04592 mus musculu |
| 26 | 78.5 | 5.9 | 442 | 1 CHMO AMATR | Q031e1 anarantus |
| 27 | 78 | 5.8 | 669 | 1 MYBE AVILE | P01105 avian leuko |
| 28 | 77.5 | 5.8 | 1202 | 1 JAG2 EAT | P97607 rattus norv |
| 29 | 77.5 | 5.8 | 1328 | 1 AGR1 DISOM | Q50404 discopysc o |
| 30 | 77.5 | 5.8 | 1391 | 1 N155 HUMAN | O75694 homo sapien |
| 31 | 77 | 5.7 | 366 | 1 CAS4 EPHMC | P18503 ephydatia m |
| 32 | 77 | 5.7 | 369 | 1 DNAB NITEU | O06431 nitrosomona |
| 33 | 77 | 5.7 | 1639 | 1 LMGI1 DROME | P15215 drosophila |

| | | | | | |
|----|------|-----|------|--------------|--------------------|
| 34 | 76.5 | 5.7 | 155 | 1 HOPD SALTY | O68927 salmonella |
| 35 | 76.5 | 5.7 | 680 | 1 CA1A HUMAN | Q03692 homo sapien |
| 36 | 76.5 | 5.7 | 770 | 1 PK7 MOUSE | Q01139 mus musculu |
| 37 | 76.5 | 5.7 | 783 | 1 PK7 RAT | Q62849 rattus norv |
| 38 | 76.5 | 5.7 | 5703 | 1 MU5B HUMAN | Q9hc84 homo sapien |
| 39 | 76 | 5.7 | 379 | 1 DNAB LEGPN | P50025 legionella |
| 40 | 76 | 5.7 | 617 | 1 SPH2 MOUSE | Q9jia7 mus musculu |
| 41 | 76 | 5.7 | 633 | 1 MUTL PSBAE | Q9hul6 pseudomonas |
| 42 | 76 | 5.7 | 674 | 1 CA1A CHICK | P08125 gallus gall |
| 43 | 75.5 | 5.6 | 785 | 1 PK7 HUMAN | Q16549 homo sapien |
| 44 | 75.5 | 5.6 | 1411 | 1 Y297 HUMAN | O15040 homo sapien |
| 45 | 75.5 | 5.6 | 1959 | 1 AGRI RAT | P25304 rattus norv |

ALIGNMENTS

RESULT 1
CA34 HUMAN
ID CA34 HUMAN STANDARD; PRT; 1670 AA.
AC Q01955; Q9BQT2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364394; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. Biol. Chem. 269:23013-23017 (1994).
RN [2]
RP REVISIONS.
RA Leinonen A.;
RN [3]
SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064596; PubMed=1134255;
RA Heidet L., Arrondelet C., Forestier L., Cohen-Solal L., Mollet G.,
Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106 (2001).
RN [4]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:19780-19784 (1992).
RN [5]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1882840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554 (1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RX TISSUE=Kidney;
MEDLINE=92147878; PubMed=1737849;


```

Query Match      100.0%; Score 1340; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 1.4e-116;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKKGKDGSGSPATWTRGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNORAHQD 60
DB 1427 GLKKGKDGSGSPATWTRGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNORAHQD 1486
QY 61 LGTLGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAPITGRALPEYIS 120
DB 1487 LGTLGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAPITGRALPEYIS 1546
QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQALASPGSCLE 180
DB 1547 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQALASPGSCLE 1606
QY 181 EFRASPLECHGRGTCNYNSYSFWLASLNPERMFKPIPTSVKAGELEKIIISRCQVCM 240
DB 1607 EFRASPLECHGRGTCNYNSYSFWLASLNPERMFKPIPTSVKAGELEKIIISRCQVCM 1666
QY 241 KGRH 244
DB 1667 KGRH 1670

RESULT 2
CA34_BOVIN
ID CA34_BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OC Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Redders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RT encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=89330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Buckowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire',

```

```

meshwork together with laminins, proteoglycans and entactin/
nidogen.
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
-!- SUBCELLULAR LOCATION: Cell surface (Potential).
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
-!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63139; AB062708.1; -.
CC FIR; A39024; A39024.
CC InterPro; IPR008150; Collagen.
CC InterPro; IPR001442; Procollagen4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 4.
CC ProDom; PD003923; ProcollagenC4; 1.
CC SMART; SM00111; C4; 2.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT SITE 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47595 MW; C03B6F14E7008DE CRC64;

Query Match      90.3%; Score 1210.5; DB 1; Length 471;
Best Local Similarity 90.6%; Pred. No. 3.7e-105;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKKGKDGSGSPATWTT-RGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNORAHQ 59
DB 227 GLKKGKDGSGSPATWTT-RGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNORAHQ 286
QY 60 DLGTGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAPITGRALPEYI 119
DB 287 DLGTGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAPITGRALPEYI 346
QY 120 SRCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQALASPGSCLE 179
DB 347 SRCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQALASPGSCLE 406
QY 180 EFRASPLECHGRGTCNYNSYSFWLASLNPERMFKPIPTSVKAGELEKIIISRCQVC 239
DB 407 EFRASPLECHGRGTCNYNSYSFWLASLNPERMFKPIPTSVKAGELEKIIISRCQVC 466
QY 240 MKRH 243

```


DR EMBL; M25561; AAAS3098.1; JOINED.
 DR EMBL; M25562; AAAS3098.1; JOINED.
 DR EMBL; M25563; AAAS3098.1; JOINED.
 DR EMBL; M25564; AAAS3098.1; JOINED.
 DR EMBL; M25565; AAAS3098.1; JOINED.
 DR EMBL; M25566; AAAS3098.1; JOINED.
 DR EMBL; M25567; AAAS3098.1; JOINED.
 DR EMBL; M25568; AAAS3098.1; JOINED.
 DR EMBL; M25569; AAAS3098.1; JOINED.
 DR EMBL; M25570; AAAS3098.1; JOINED.
 DR EMBL; M25571; AAAS3098.1; JOINED.
 DR EMBL; M25572; AAAS3098.1; JOINED.
 DR EMBL; M25573; AAAS3098.1; JOINED.
 DR EMBL; M25574; AAAS3098.1; JOINED.
 DR EMBL; M25575; AAAS3098.1; JOINED.
 DR EMBL; Y00706; CAA68698.1; -.
 DR EMBL; X05561; CAA29075.1; -.
 DR EMBL; M10940; AAAS2006.1; -.
 DR EMBL; M11315; AAAS2042.1; -.
 DR PIR; S16876; CGHU4B.
 DR Genew; HGNC:2202; COL4A1.
 DR MIM; 120130; -.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD000007; Collagen; 24.
 DR ProDom; PD003923; Procollagn4; 1.
 DR SMART; SM00111; C4; 2.
 DR Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT CARBOHYD 126 126 NONHELICAL REGION (NC1).
 FT DISULFID 1460 1551 N-LINKED (GLCNAC. . .).
 FT DISULFID 1493 1548 OR 1548.
 FT DISULFID 1505 1511 OR 1551.
 FT DISULFID 1570 1665 OR 1662.
 FT DISULFID 1604 1662 OR 1665.
 FT DISULFID 1616 1622
 FT CONFLICT 237 238 SG -> K (IN REF. 4).
 FT CONFLICT 241 241 G -> K (IN REF. 4).
 FT CONFLICT 319 319 Q -> A (IN REF. 3).
 FT CONFLICT 719 719 N -> D (IN REF. 5).
 FT CONFLICT 837 837 D -> Y (IN REF. 5).
 FT CONFLICT 842 842 K -> P (IN REF. 5).
 FT CONFLICT 896 896 V -> W (IN REF. 2).
 FT CONFLICT 904 904 E -> Q (IN REF. 5).
 FT CONFLICT 914 914 S -> K (IN REF. 5).
 FT CONFLICT 998 998 S -> K (IN REF. 5).
 FT CONFLICT 1010 1010 X -> P (IN REF. 5).
 FT CONFLICT 1012 1012 E -> K (IN REF. 5).
 FT CONFLICT 1358 1358 E -> K (IN REF. 5).
 FT SEQUENCE 1669 AA; 160611 MW; 3BBA6DFF9B9A84 CRC64;
 Query Match
 Best Local Similarity 71.6%; Score 960; DB 1; Length 1669;
 Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;
 QY 1 GLKGRGDSGPATWTRGTFVTRHSQTALPSCEGTVPYSGFSFLVQGNQRAHQD 60
 DB 1429 GLPGSMGPPGTPS--VDHGFLVTRHSQTIDDPQCPSTKILYHGYSLLVYQGNRAHQD 1486
 QY 61 LGTGLSCLORTTMBPLFCNVNDYCNFASRNDYSYWLSTPALMPMNAIPITGRALPEYIS 120
 DB 1487 LGTAGSCURKSTMPPLFCNVNDYCNFASRNDYSYWLSTPEPMPMNAIPITGENIRDFIS 1546

QY 121 RCTVCEGPAIAIAVHSOTTDIPCPHGWIISLWKGFSFIMFTSAGSEGTGOALASPGSCLE 180
 DB 1547 RCACEAPAMVAVHSQTIQIPPCSGWSSLIWIGISFVWHTSAGBSGQALASPGSCLE 1606
 QY 181 EFRASPFLCEHGRGTCNYNSYSFSLASLNPFRKPIPTSTYKAGELEKIIISRCQVCM 240
 DB 1607 EFRSAPFTECHGRGTCNYANAYSWFLATIRSEMFKKPTSTLKAGELRTHVSRQVCM 1666
 QY 241 KK 242
 DB 1667 RR 1668
 RESULT 4
 CA14_MOUSE
 ID CA14_MOUSE STANDARD; PRT; 1669 AA.
 AC F02453;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89197932; PubMed=2703490;
 RA Muthukumar G., Blumberg B., Kurkinen M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 collagen IV. Differential evolution of collagen IV domains.";
 RL J. Biol. Chem. 264:6310-6317(1989).
 RN [2]
 RP SEQUENCE OF 1-1154 FROM N.A.
 RX MEDLINE=88112221; PubMed=3338568;
 RA Wood L., Theriault N., Vogeli G.;
 RT "cDNA clones completing the nucleotide and derived amino acid
 sequence of the alpha 1 chain of basement membrane (type IV) collagen
 from mouse.";
 RL FEBS Lett. 227:5-8(1988).
 RN [3]
 RP SEQUENCE OF 1149-1424 FROM N.A.
 RX MEDLINE=86301886; PubMed=3755692;
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
 synthetic oligodeoxynucleotide.";
 RL Gene 43:301-304(1986).
 RN [4]
 RP SEQUENCE OF 1276-1669 FROM N.A.
 RX MEDLINE=85127033; PubMed=2578961;
 RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
 RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
 the alpha 1(IV) chain of basement membrane collagen as derived from
 complementary DNA.";
 RL Eur. J. Biochem. 147:217-224(1985).
 RN [5]
 RP SEQUENCE OF 1441-1669 FROM N.A.
 RX MEDLINE=87250460; PubMed=3597383;
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
 RA Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse
 alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8496-8499(1987).
 RN [6]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=86196099; PubMed=3009468;
 RA Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
 collagen genes.";
 RL J. Biol. Chem. 261:6654-6657(1986).
 RN [7]

RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen
RT chain and identification of a single-base mutation in exon 23
RT converting glycine 541 in the collagenous domain to cysteine in an
RT Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RN SEQUENCE OF 85-1685 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=90337990; PubMed=2380186;
RX Pihlajaniemi T., Pohjola-Evonen E.R., Myers J.C.;
RA "Complete primary structure of the triple-helical region and the
RT carboxyl-terminal domain of a new type IV collagen chain, alpha
RT 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RN SEQUENCE OF 924-1685 FROM N.A.
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha
RT 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
RN [5]
RN SEQUENCE OF 914-1685 FROM N.A.
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
RA Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with
RT restricted kidney distribution and assignment of its gene to the
RT locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RN SEQUENCE OF 1442-1471 FROM N.A.
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pihlajaniemi E.R., Kadri A.S., Goddard A.D.,
RA Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
RT to the region of the X chromosome containing the Alport syndrome
RT locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RN SEQUENCE OF 1-20 FROM N.A.
RP SEQUENCE OF 1-20 FROM N.A.
RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
RA Marynen P.;
RT Submitted (SEP-1994) to the ENBL/GenBank/DBJ databases.
RN [8]
RN SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
RA Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood
RT cells: a complex mutation in the COL4A5 gene of an Alport patient
RT deletes the NCI domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RN REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RN VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
RA Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
RT conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RN VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
RA Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5
RT (IV) chain associated with X-linked Alport syndrome: characterization
RT of the mutation by direct sequencing of PCR-amplified lymphoblast
RT cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RN VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
RA Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RN VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of
RT patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
RN [14]
RN VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;
RP ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
RN [15]
RN VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
RN [16]
RN VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zarelli P., Nexi T.M., Rossetti S.,
RA Turco A.E., Heiskari M., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Seesa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
RN [17]
RN VARIANTS AS, AND VARIANTS ASP-430; SER-444; ASN-664 AND
RP MET-1428.
RX MEDLINE=97094179; PubMed=8940267;
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
RN [18]
RN VARIANT AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobrykina V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of alpha 5(IV) collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
RN [19]
RN VARIANT AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).

[20]
RN VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
Pignatti G.F., Galli L., Pruttini M., Renieri A., Mingarelli R.,
Tivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.,
"Wissense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1: S106-S109 (1998).
[21]
RN VARIANTS AS VAL-420; 456; PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP ARG-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
Sumino K., Nishiyama K., Iijima K., Yoshikawa N.,
"Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
sequencing.";
RL Am. J. Kidney Dis. 34: 854-862 (1999).
[22]
RN VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
Sumino K., Nishiyama K., Iijima K., Yoshikawa N.,
"Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
sequencing.";
RL Am. J. Kidney Dis. 34: 854-862 (1999).
[23]
RN VARIANTS AS ARG-822.

Query Match 70.4%; Score 943.5; DB 1; Length 1685;
Best Local Similarity 67.9%; Pred. No. 1.2e-79;
Matches 165; Conservative 35; Mismatches 42; Indels 1; Gaps 1;
QY 1 GLKGRKDGSGSPATWT-TRGFVFTRHSTQTAIPSCPGTGLVSGFSLVQGNRAHQ 59
DB 1442 GPDGLQGPFGPTGSSVAHGFILTRHSQTTDAPCQPTGLVQVGFSLVQGNRAHQ 1501
QY 60 DLGTLGSLQFTTMTPLFCNVNDVCFNSNDYSYWLSTPALPMNMAITGEALPYI 119
DB 1502 DLGTAGSLRFRSTPMFNCNNVCFNSNDYSYWLSTPEPMNSQPLKGSQTQFFI 1561
QY 120 SRCTVCEGPAIAVHSQTTDIPCPHGWISLWGFSGFIMFTSAGSGCTGQALASPGSCL 179
DB 1562 SRCVCEAPAVVIAVHSQTTIQCPCQGWDSLWIGYFPMHTSAGSGQALASPGSCL 1621
QY 180 EFRASPFLCHGRTGTCNYNSYSFWLASLNPMPFKPIPTSVKAGELEKILSRQVC 239
DB 1622 EFRSAPFLCHGRTGTCNYNSYSFWLATVDVSDMFSPKQSETIKAGDLRTISRQVC 1681
QY 240 MKK 242
DB 1682 MKR 1684

RESULT 6
ID CA54 CANFA STANDARD; PRT; 754 AA.
AC Q28247.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=Sanoyed; TISSUE=Kidney;
RX MEDLINE=9424866; PubMed=8171024;
RA Zheng K., Thorne P.S., Marrano P., Bauml R., McInnes R.R.;
"Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993 (1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
CC canine X-linked hereditary nephritis (HN), a disease similar to
CC that in humans (also referred to as Alport syndrome) characterized
CC by progressive renal failure and neurosensory deafness.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U07888; AAB50258.1; -.
DR PIR; A55267; A55267. Clg helix.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; Clg helix; 1.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON TER 1 1
FT DOMAIN <1 530 TRIPLE-HELICAL REGION.
FT DOMAIN 531 >754 NONHELICAL REGION (NC1).
FT DISULFID 552 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 ? BY SIMILARITY.
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON TER 754 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 68.8%; Score 922.5; DB 1; Length 754;
Best Local Similarity 68.4%; Pred. No. 4.1e-78;
Matches 162; Conservative 32; Mismatches 42; Indels 1; Gaps 1;

QY 1 GLKGRKDGSGSPATWT-TRGFVFTRHSTQTAIPSCPGTGLVSGFSLVQGNRAHQ 59
DB 518 GPDGMQGPFGPTGSSVAHGFILTRHSQTTDAPCQPTGLVQVGFSLVQGNRAHQ 577


```

Db 1635 SRCACEVPTQIIAAVHSQDTSVPCCPQSGMMTGYSFVWHTFAAGAGTGOSLQSPGSC 1694
Qy 160 EEFRAFPLECHGRTCNVYSNSYFWLASLNMERFRKPISTVKAGLELKIISRCQVC 239
Db 1695 EEFRAVPIECHGRTCNVYATNHGFWSPIVDQDKQFRKEMSGTLKAGGLKDRVSRQVC 1754
Qy 240 MKKR 243
Db 1755 LKNR 1758
RESULT 8
CR24_ACSU
ID CA24_ACSU STANDARD; PRT; 1763 AA.
AC P27393;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
ID_1 RN
SEQUENCE FROM N.A. (ISOFORMS I AND II).
RP MEDLINE=91340768; PubMed=1714907;
RX Pettitt J., Kingston I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
RT and the partial structural organization of its gene.";
RL J. Biol. Chem. 266:16149-16156(1991).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I;
CC IsoId=P27393-1; Sequence=Displayed;
CC Name=II;
CC IsoId=P27393-2; Sequence=VSP 001159;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; M67507; AA18014.1; -.
DR PIR; S16366; S16366.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollag_n_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD000323; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;

```

genes are located on separate chromosomes.";
 [3] J. Biol. Chem. 264:17574-17582(1989).
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=Br1stol N2;
 RA Wu X., Le T.T.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 EN [4]
 RP VARIANTS.
 RX MEDLINE=94320591; PubMed=80452559;
 RA Sibley M.H., Graham P.L., von Mendel N., Kramer J.M.;
 RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
 Ctenorhynchus elegans produce phenotypes of differing severities";
 RL EMBO J. 13:3278-3285(1994).
 CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
 CC -!- Vital for embryonic development.
 CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
 CC Type IV collagen forms a mesh-like network linked through
 CC intermolecular interactions between 7S domains and between NC1
 CC domains.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=a;
 CC IsoId=P17140-1; Sequence=Displayed;
 CC Name=2; Synonyms=b;
 CC IsoId=P17140-2; Sequence=VSP 001160;
 CC -!- DEVELOPMENTAL STAGE: Isoform 1 is predominant in embryos and
 CC isoform 2 is predominant in the larvae and adults.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-
 CC X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 222964; CA880536.1; -;
 CC EMBL; 222964; CA880537.1; -;
 CC EMBL; J05066; AAA27989.1; -;
 CC EMBL; U22327; AAA64312.1; ALT_SEQ.
 CC EMBL; U53342; AAA96215.1; -;
 CC EMBL; U53342; AAA96216.1; -;
 CC PIR; T29350; T29350.
 CC WormPep; F01G12.5a; CE04334.
 CC WormPep; F01G12.5b; CE04335.
 CC GO; GO:0005587; C:collagen type IV; IMP.
 CC GO; GO:0030020; P:extracellular matrix structural constituent. .; IMP.
 CC GO; GO:0016043; P:cell organization and biogenesis; NAS.
 CC InterPro; IPR008161; Clg helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR001442; Procollagen4_C.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 23.
 CC ProbDom; PD000007; Clg helix; 6.
 CC ProbDom; PD003923; ProcollagenC4; 1.
 CC SMART; SM00111; C4; 2.
 CC Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 KW Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 1758 COLLAGEN ALPHA 2(IV) CHAIN.
 FT DOMAIN 27 42 7S DOMAIN.

FT DOMAIN 42 1527 TRIPLE-HELICAL REGION.
 FT DOMAIN 1528 1758 NON-HELICAL REGION (NC1).
 FT DISULFID 1546 1635 OR 1632 (BY SIMILARITY).
 FT DISULFID 1579 1632 OR 1635 (BY SIMILARITY).
 FT DISULFID 1591 1597 BY SIMILARITY.
 FT DISULFID 1654 1750 OR 1747 (BY SIMILARITY).
 FT DISULFID 1688 1747 OR 1750 (BY SIMILARITY).
 FT DISULFID 1700 1707 BY SIMILARITY.
 FT CARBOHYD 248 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
 FT VARSPLIC 229 GDLSGVGPDPGPPREFTGSGSIYVGNPNPKGDK -> G
 FT DIAMGAPGAPGPPGPPASTWTKGTIIIGKGLGKGEK (in
 FT isoform 1).
 FT /FTID-VSP 001160.
 FT G -> E (IN MN114; 73% LETHAL).
 FT VARIANT 48
 FT G -> T (IN MN126; 100% LETHAL).
 FT VARIANT 366
 FT G -> E (IN MN109; 37% LETHAL).
 FT VARIANT 570
 FT G -> R (IN MN103 AND MN151; 96% LETHAL).
 FT VARIANT 588
 FT G -> R (IN MN152; 50% LETHAL).
 FT VARIANT 597
 FT G -> R (IN MN101; 100% LETHAL).
 FT VARIANT 690
 FT G -> E (IN MN129; 100% LETHAL).
 FT VARIANT 690
 FT G -> E (IN MN143; 100% LETHAL).
 FT VARIANT 737
 FT G -> R (IN G30; 90% LETHAL).
 FT VARIANT 877
 FT G -> R (IN E1470; 94% LETHAL).
 FT VARIANT 904
 FT VARIANT 1003
 FT VARIANT 1125
 FT G -> E (IN MN139; 20% LETHAL).
 FT VARIANT 1152
 FT G -> D (IN MN147; 7% LETHAL).
 FT VARIANT 1286
 FT G -> D (IN G37 AND B246; 9% LETHAL).
 FT CONFLICT 1604
 FT CONFLICT 1682 P -> L (IN REF. 1 AND 3; AAA96216).
 FT SEQUENCE 1758 AA; 167750 MW; 97E3F3DBB2D2AC5 CRC64;
 Query Match 57.6%; Score 772.5; DB 1; Length 1758;
 Best Local Similarity 54.5%; Pred. No. 1e-63;
 Matches 140; Conservative 42; Mismatches 60; Indels 15; Gaps 4;
 QY 1 GLKKGKDGSGSP-----ATWTRGVFTRHQTTPAISCPRGTVPVLYSGRSEFLVQ 51
 DB 1504 GLDGQPGGAGLPGAGPAGVADGFLVKHSQTTEVPCRPGEQTKLMDGYSLLYIE 1563
 QY 52 GQRAHGQDLGLGSCLOQFTTMBELFCNVNDVNCNFAISNDYVWLSLTPALPMNMAPIT 111
 DB 1564 GNEKSHNDQLGHAGSCLOQFSTMPFLCDFNNVCNYSRNEKSYWLSLSEAIP--MMPVN 1621
 QY 112 GRALPPIYISRCVCGGPAIAIVHSQTTDIPCPHGWSLWKGFSPFIMFTSGSGTGOA 171
 DB 1622 ERIEPIYISRCVCGGPAIAIVHSQTTQIPNCPAGWSLWIGYSFAMHTGAGAGGGGS 1681
 QY 172 LASPGSCLEPRASPFLECHG-RGTCNYSNSYSFWLASLNPFRMFRKPISTVKAQELE 230
 DB 1682 PSSPGSCLEDFRATPFIECGARGSCHYFANKFSFWLTIDNDSFKYPEQTLKSGNLR 1741
 QY 231 KIISRCQVCMKK---RH 244
 DB 1742 TRVSRQVCVKSTDGRH 1758
 RESULT 10
 CA24_MOUSE
 ID CA24_MOUSE STANDARD; PRT; 1707 AA.
 AC P08122; Q63375;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 GN COL4A2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89197933; PubMed=2703491;
 RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,

RA Pihlajaniemi T., Kurkinen M.;
RT "The complete primary structure of mouse alpha 2(IV) collagen."
RL Alignment with mouse alpha 1(IV) collagen."
RN J. Biol. Chem. 264:6318-6324(1989).
[2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3196626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes."
RN J. Biol. Chem. 263:19274-19277(1988).
[3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=301132;
RA Schwarz U., Schuppan D., Oberbaumer I., Glanville R.W.,
RT Deutzmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RT chain and its comparison with the alpha 1(IV) chain."
RL Eur. J. Biochem. 157:49-56(1986).
[4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbaumer I., Kuehn K.;
RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1(IV)."
RN FEBS Lett. 208:203-207(1986).
[5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RT Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen."
RL J. Biol. Chem. 262:8496-8499(1987).
[6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen."
RL FEBS Lett. 206:29-32(1986).
[7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RT alpha 2(IV) collagen gene."
RL Nature 317:177-179(1985).
[8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer."
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@sib-sib.ch.
CC -----
DR EMBL; M23334; AAA51626.1; JOINED.
DR EMBL; M23333; AAA51626.1; JOINED.
DR EMBL; J04695; AAA50293.1; -
DR EMBL; J04448; AAA37438.1; -
DR EMBL; X04647; CAA28308.1; -
DR EMBL; M15833; AAA37341.1; -
DR EMBL; X04410; CAA27998.1; -
DR EMBL; X02896; CAA26655.1; -
DR EMBL; X02897; CAB51614.1; -
DR EMBL; X02898; CAA26657.1; -
DR EMBL; X02899; CAA26658.1; -
DR PIR; A33526; A33526.
DR MGI; MGI:88455; Col4a2.
DR CO; CO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Cig_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; P0000007; Cig_helix; 7.
DR ProDom; P0003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.
FT CHAIN 184 1707 TRIPLE-HELICAL REGION.
FT DOMAIN 184 1479 NONHELICAL REGION (NC1).
FT DISULFID 1499 1598 OR 1585 (BY SIMILARITY).
FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).
FT DISULFID 1544 1550 BY SIMILARITY.
FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).
FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).
FT DISULFID 1653 1660 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1270 1270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1051 1051 P -> R (IN REF. 6).
FT CONFLICT 1097 1097 S -> G (IN REF. 7).
FT CONFLICT 1171 1171 G -> S (IN REF. 6).
FT CONFLICT 1179 1179 P -> R (IN REF. 6).
FT CONFLICT 1241 1241 Q -> E (IN REF. 6).
FT CONFLICT 1328 1328 P -> A (IN REF. 6).
FT CONFLICT 1573 1573 V -> L (IN REF. 4).
FT CONFLICT 1623 1623 Y -> H (IN REF. 4).
SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;
Query Match 56.8%; Score 760.5; DB 1; Length 1707;
Best local similarity 57.6%; Pred. No. 1.3e-62;
Matches 140; Conservative 37; Mismatches 61; Indels 5; Gaps 4;
QY 1 GLKRGKSGSPATWTRGFVTRHSQTALIPSCPEGTVPVLYSGFSLFVQGNRAHQD 60
1466 GRGSPGLFGMPGRSVISGYLLVSKHSQTDQEPMPVGMNKLWSGYSLLYFEGQEKAHQD 1525
DB 61 LGTLGSLQRTFTMPFLPCNVNDVNCNFRNDYSWLSTPALMPNMAPIGTRALEPVIS 120
1526 LGLAGSLARFTMPFLYCNPDVCYASRNDKSWLSTTA--PLPMPVAAEEIKPVIS 1583
QY 121 RCTVCEGPAIAVAHSQTDDIPCPHGMIISLWKGSFTMTFSAGSEGTQALASPGSCL 180
1584 RSVCEAPAVAVHSQDTSIHCPAGWRSLSWISFLMYTAGDEGGQSIVSPGSCLE 1643
QY 181 EFRASPFLFLECH-GRGTCNTYNSYSFSLASLNPFRFK-PIPTVTKAGELEKIISRCQV 238

Db 1644 DFRATPEICNGRGTCYFANKYSFWLTTI-PEQNFQSPSADTLKAGLIRTHSRCQV 1702
QY 239 CMK 241
Db 1703 CMK 1705
RESULT 11
CA24 HUMAN
ID CA24 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
collagen and comparison with the alpha 1(IV) chain.";
RL J. Biol. Chem. 263:19488-19493(1988).
RN [2]
RP SEQUENCE OF 1-1042 FROM N.A.
RX MEDLINE=88151998; PubMed=3345760;
RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;
RT "Human basement membrane collagen (type IV). The amino acid sequence
of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
chain reveals deletions in the alpha 1(IV) chain.";
RL Eur. J. Biochem. 172:35-42(1988).
RN [3]
RP SEQUENCE OF 1254-1712 FROM N.A.
RX MEDLINE=87219158; PubMed=3582677;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2
chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
(IV) but not with the collagenous domain or 3'-untranslated region.";
RL FEBS Lett. 216:281-286(1987).
RN [4]
RP SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-515(1987).
RN [5]
RP SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-
specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
RL J. Biol. Chem. 262:9231-9238(1987).
RN [6]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soiminen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
collagen are divergently encoded on opposite DNA strands and have an
overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [7]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89030632; PubMed=2846280;
RA Poeschl E., Pollner R., Kuehn K.;
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
basement membrane collagen type IV are arranged head-to-head and

separated by a bidirectional promoter of unique structure.";
RL EMBO J. 7:2687-2695(1988).
RN [8]
RP SEQUENCE OF 1-33 FROM N.A.
RX TISSUE=Skin;
RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
RT "Identification of a novel sequence element in the common promoter of
region of human collagen type IV genes, involved in the regulation of
divergent transcription.";
RL Biochem. J. 292:687-695(1993).
RN [9]
RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RX TISSUE=Placenta;
RA Siebold B., Deutmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
carboxyterminal, non-collagenous aggregation and cross-linking domain
of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
RN [10]
RP FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
or send an email to license@ebi-sib.ch).
CC
CC EMBL; X05562; CAA29076.1; -
CC EMBL; X05610; CAA29098.1; -
CC EMBL; J02760; AAA58422.1; -
CC EMBL; M36963; AAA53099.1; -
CC EMBL; X12784; CAA31275.1; -
CC EMBL; J04217; AAA53097.1; -
CC PIR; A32024; CGHU2B.
CC Genew; HGNC:2203; COL4A2.
CC MIM; 120090; -
CC GO; GO:0005587; C:collagen type IV; TAS.
CC GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
CC GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
CC InterPro; IPR008161; Clg helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagn4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 24.
CC ProDom; PD000007; Clg helix; 7.
CC ProDom; PD003923; ProcollagnC4; 1.
CC SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25
FT PROPEP 26 183
FT CHAIN 184 1712
FT DOMAIN 184 1484

```
FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
FT DISULFID 1549 1555 BY SIMILARITY.
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
FT DISULFID 1658 1665 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
FT CONFLICT 471 471 R -> P (IN REF. 2).
FT CONFLICT 583 583 A -> I (IN REF. 2).
FT CONFLICT 1575 1575 M -> I (IN REF. 5).
FT CONFLICT 1563 1663 G -> H (IN REF. 9).
FT CONFLICT 1701 1701 H -> G (IN REF. 9).
SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match 56.4%; Score 755.5; DB 1; Length 1712;
Best Local Similarity 58.08; Pred. No. 3.8e-62;
Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;

Qy 1 GLKGRGDSGPATWTRGFTVTRHSQTTPAIPCEGTVPYSGFSLFVQGNQRAHQD 60
Db 1471 GRPGSPGLPGMGSRVSGVLLVKHSQTDQBPMPGVGMNKLWSGYSLLYFEGQEKAHNQD 1530
Qy 61 LGTIGSCLOFTTTPFFLCNVNDVNCNPNASNDYSLWSTALMPNMAPITGREALPEYIS 120
Db 1531 LGLAGSCLARFSTPFPYCNPDGVCYASRNDKSYWLSTTA--PLPMPVPAEDEIKPYIS 1588
Qy 121 RCTVCEGPALAIHVSTQTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
Db 1589 RCVCEAPALAIHVSDVSIHPCHPAGWRSIMIGYSLFMTAAGDEGGQSLVSPGSCLE 1648
Qy 181 EFRASPLECH-GRGTQNYNSNSVFWLASINBERMPR-KPISTVKGAELEKILRCQV 238
Db 1649 DFRATPFIECNGRGRTCHYANKYSFWLTTI-PEQSPQGSFSDTLKAGLIRTHIRGQV 1707
Qy 239 CMK 241
Db 1708 CMK 1710

RESULT 12
CA64 HUMAN STANDARD; PROT: 1691 AA.
AC Q14031; Q12823; Q14053; Q9NQMS; Q9NTX3; Q9UJ76; Q9UMG6; Q9Y4L4;
DI 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RX MEDLINE=94171779; PubMed=8125972;
RA Ohashi T., Sugimoto M., Watanabe M.-G., Ninomiya Y.;
RT Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.;
RL J. Biol. Chem. 269:7520-7526(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Reders S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
```

```
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Reders S.T., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grahame D., Lawlor S., Wilson S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=8356449;
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,
RA de Paep A., Tryggvason K., Reders S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169(1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D21337; BAA04809.1; -
CC EMBL; U04845; AAA19569.2; -
CC EMBL; U47004; AAB19038.1; -
CC EMBL; U46959; AAB19038.1; JOINED.
CC EMBL; U46961; AAB19038.1; JOINED.
CC EMBL; U46962; AAB19038.1; JOINED.
CC EMBL; U46963; AAB19038.1; JOINED.
CC EMBL; U46964; AAB19038.1; JOINED.
CC EMBL; U46965; AAB19038.1; JOINED.
CC EMBL; U46966; AAB19038.1; JOINED.
CC EMBL; U46967; AAB19038.1; JOINED.
CC EMBL; U46968; AAB19038.1; JOINED.
CC EMBL; U46969; AAB19038.1; JOINED.
CC EMBL; U46970; AAB19038.1; JOINED.
CC EMBL; U46971; AAB19038.1; JOINED.
CC EMBL; U46972; AAB19038.1; JOINED.
CC EMBL; U46973; AAB19038.1; JOINED.
CC EMBL; U46974; AAB19038.1; JOINED.
CC EMBL; U46975; AAB19038.1; JOINED.
CC EMBL; U46976; AAB19038.1; JOINED.
CC EMBL; U46977; AAB19038.1; JOINED.
```

| | |
|--|--|
| DR | ENBL; ALI36080; CAB96748.1; -. |
| DR | ENBL; ALO031177; CAA20120.1; -. |
| DR | ENBL; L22763; AAA16338.1; -. |
| DR | PIR; A54122; CGHUEB. |
| DR | Genew; HGNC:2208; COL4A6. |
| DR | MIM; 303631; -. |
| DR | GO; GO:0005587; C:collagen type IV; NAS. |
| DR | GO; GO:0005201; Extracellular matrix structural constituent; NAS. |
| DR | GO; GO:0030198; Extracellular matrix organization and bioge. .; NAS. |
| DR | InterPro; IPR008161; Clg_helix. |
| DR | InterPro; IPR008160; Collagen. |
| DR | InterPro; IPR001442; Procollagn_4_C. |
| DR | Pfam; PF01413; C4; 2. |
| DR | Pfam; PF01391; Collagen; 23. |
| DR | ProDom; PD000007; Clg_helix; 4. |
| DR | ProDom; PD003923; ProcollagnC4; 1. |
| DR | SMART; SM00111; C4; 2. |
| KW | Extracellular matrix; Connective tissue; Basement membrane; |
| KW | Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Signal; |
| KW | Alternative splicing; Polymorphism. |
| FT | SIGNAL 1 22 POTENTIAL. |
| FT | CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN. |
| FT | DOMAIN 23 46 7S DOMAIN. |
| | |
| Qy | Query Match 55.1%; Score 739; DB 1; Length 1691; |
| Db | Best Local Similarity 54.7%; Pred. No. 1.3e-60; |
| Db | Matches 133; Conservative 42; Mismatches 64; Indels 4; Gaps 3; |
| | |
| Qy | 1 GLKGRDSSPATWTTRGFVFRHSQTATPSCPCTGTPLYSGFSLFYQGNGRAHQD 60 |
| Db | 1449 GQQGPFMPGMPGQSMRVGYTLVKHSQEQVPCPIGMSQLVGYSLLFVEGGKXHNQD 1508 |
| | |
| Qy | 61 LGTIGSCLOFTTMPFLFCNVNDVCNFAISNDYSYWLSTPALPMNNAPITGRALEFYIS 120 |
| Db | 1509 LGFAGSCLPFSFTWFYICNINEVCHVARENDKSXYLSTTA--PIPMPVSQTQIPQYIS 1566 |
| | |
| Qy | 121 RCTVCEGPATAIAVHSOTTDIPCPHGWSLWGFGFIMFTSAGSEGTGAALASPGSCLE 180 |
| Db | 1567 RCSYCEAPSQAIAVHSODTIPOCPGLWRSLMTGYSLFMHTAAGAEGGOSLVSPGSCLE 1626 |
| | |
| Qy | 181 EFRASPPECWG-RGTCNNYSNSYSFWLASLNPERMFK--PIPTKVKALEIKIIIRCQV 238 |
| Db | 1627 DFRATPIECSGARGICHYFANKYSFWLTVERQFGELPVSETLKAGQLHTEVSRQV 1686 |
| | |
| Qy | 239 CMK 241 |
| Db | 1687 CMK 1689 |
| | |
| RESULT 13 | |
| CA44_BOVIN | STANDARD; PRT; 453 AA. |
| ID CA44_BOVIN | |
| AC Q29442; | |
| DT 01-NOV-1997 (Rel. 35, Created) | |
| DT 01-NOV-1997 (Rel. 35, Last sequence update) | |
| DT 15-MAR-2004 (Rel. 43, Last annotation update) | |
| DE Collagen alpha 4(IV) chain (Fragment). | |
| GN COL4A4. | |
| OS Bos taurus (Bovine). | |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; | |
| OC Bovidae; Bovinae; Bos. | |
| OX NCBI_TaxID=9913; | |
| RN [1] | |
| RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328. | |
| RC TISSUE=Lens; | |
| RR MEDLINE=92112769; PubMed=1370461; | |
| RA Mariyama M., Kalluri R., Hudson B.G., Reenders S.T.; | |
| RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of | |
| RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV | |
| RL collagens."; | |
| RL J. Biol. Chem. 267:1253-1258(1992). | |
| RN [2] | |

SEQUENCE OF 217-246.
MEDLINE=90202779; PubMed=2318822;
Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
"Glomerular basement membranes. Identification of a fourth chain,
alpha 4, of type IV collagen."
J. Biol. Chem. 265:5466-5469(1990).
[3]
SEQUENCE OF 217-233.
MEDLINE=87222419; PubMed=2438283;
Buckowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
Hudson B.G.;
"Localization of the Goodpasture epitope to a novel chain of basement
membrane collagen."
J. Biol. Chem. 262:7874-7877(1987).
-!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
-!- SUBCELLULAR LOCATION: Cell surface (Potential).
-!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
colocalized and present only in basement membranes of kidney, eye,
cochlea, lung and brain.
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
-!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M77480; AAA30458.2; ALT_SEQ.
PIR; S18804; S18804.
InterPro; IPR008160; Collagen.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 4.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1
FT DOMAIN <1 222 TRIPLE-HELICAL REGION.
FT DOMAIN 223 453 NONHELICAL REGION (NC1).
FT DISULFID 243 332 OR 329 (BY SIMILARITY).
FT DISULFID 276 329 OR 332 (BY SIMILARITY).
FT DISULFID 288 294 BY SIMILARITY.
FT DISULFID 351 449 OR 446 (BY SIMILARITY).
FT DISULFID 385 446 OR 449 (BY SIMILARITY).
FT DISULFID 397 404 BY SIMILARITY.
FT CONFLICT 219 219 I -> P (IN REF. 2 AND 3).
SQ SEQUENCE 453 AA; 46384 MW; FTED40AE9A65BC1 CRC64;
Query Match 52.9%; Score 709; DB 1; Length 453;
Best Local Similarity 49.8%; Pred. No. 1.7e-58;
Matches 130; Conservative 42; Mismatches 67; Indels 22; Gaps 5;
QY 1 GLKGRGDSGSPAT-----WTRGEVFRHSQTATPSCPSTPLYS 43

Db 193 GHKDMGEACGCGAGPPGMDGPIGFGYLSGFLVLVHSQTGDEPTCPMGMLWT 252
QY 44 GFSFLVQGNQRAHGQDLGTLSCLQRFTTNPFLFCNVNDVCFASRNDYSYLSLTPALM 103
Db 253 GYSLYLEGQBRANHQDLGSLPTIFSLTPAYCNHQVCHYARRNDSYWLASTA-- 310
QY 104 PMNMAPITGRALBPIYSRCTVCEGPAIAIAVHSOTTDIPPCPGHWSLWKGFIFMFTA 163
Db 311 PLMTPLSEDEIRPIYSRCAVCAQAQAVAHSDQSIPECPRAWRLWTGYSLMHTGA 370
QY 164 GSEGTGALASPGSCLEEFASPLECHGR-GTCNYYSNSYFWLASLNLNPERMPK-PIP 221
Db 371 GDQGGGALASPGSCLEDFEAFAPLECGQRGTCHFFANKYSFWLITVRDLOFSSAPLP 430
QY 222 STVKAQELK-IISRCQVCMK 241
Db 431 DTLKSHAQROKISRCQVCVK 451
RESULT 14
CA44_RABIT
ID CA44_RABIT STANDARD; PRT; 623 AA.
AC P55787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4 (IV) chain (Fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
alpha 4 chain of basement membrane collagen type IV and assignment of
the gene to the distal long arm of human chromosome 2.";
J. Biol. Chem. 267:23753-23758(1992).
RL
CC -!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the G-
X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L01477; -! NOT_ANNOTATED_CDS.
PIR; A45137; A45137.


```
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SW00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT DOMAIN 393 623 NON-HELICAL REGION (NC1).
FT DISULFID 413 502 OR 499 (BY SIMILARITY).
FT DISULFID 446 499 OR 502 (BY SIMILARITY).
FT DISULFID 458 464 BY SIMILARITY.
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 OR 619 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SQ SEQUENCE 623 AA; 62393 MW; CCB9BB31242FE92 CRC64;

Query Match 52.4%; Score 702; DB 1; Length 623;
Best Local Similarity 49.0%; Pred. No. 1.1e-57;
Matches 128; Conservative 45; Mismatches 56; Indels 22; Gaps 5;

Qy 1 GLKGRGDSGPAT-----WTRGFVFRHSQTVAIPSCPGTVPLYS 43
Db 363 GHKGDTCGEAGRPGAPGPPPTGDPGKGLGPGYLSGFLVLSQTDQEPACPMGPRLWT 422
Qy 44 GFSFLFVQGNORAGQDLGLTGLSCLOREFTTMRPFLFCNVNVDVCFNDRNDYSYWLSTPALM 103
Db 423 GYSLLYEGEKEKANKQDLGLAGSLPFTSLPFCYCNHCVHQAQRNDKSYWLASAG-- 480
Qy 104 PMNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIPPCPGWISLWKGFSFIMFTSA 163
Db 481 PLPMWPLSEBIRYISKCAVCEAPAAQAVHSDQSIQPCPRAWRLWISLWKGFSFIMFTSA 540
Qy 164 GSEGTQALASPGSCLEEFASPLERCHGR-GTCNYNSYNSYWLASLAPD-RMFRKPIP 221
Db 541 GDQGGQALWSPGSCLEDFEAPLECGRQGTCHFFANYSFWLTVTPDQLQVFSAPSP 600
Qy 222 STVRAGELEK-IISRCQVCMK 241
Db 601 DTLKESQAQRKISRCQVCVK 621

RESULT 15
CA44 HUMAN
ID CA44 HUMAN STANDARD; PRT; 1690 AA.
AC P53420.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 4(IV) chain precursor.
GN COL4A4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxonomy:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=95014445; PubMed=7523402;
RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reenders S.T.;
RT "Complete primary structure of the human type IV collagen alpha 4(IV)
RT chain. Comparison with structure and expression of the other alpha
RT (IV) chains.";
RL J. Biol. Chem. 269:26172-26177(1994).
RN [2]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Moneta R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
```


GenCore version 5.1.6
Copyright (C) 1993 - 2004: CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 146.518 Seconds
(without alignments)
525.440 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 1340

Sequence: 1 GLKGRGDSGSPATWTRGF.....KAGELEKIISRCQVQMKKRH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mmc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|---------------------|
| 1 | 1340 | 100.0 | 245 | 4 Q9NYC4 | Q9nyc4 homo sapien |
| 2 | 1221.5 | 91.2 | 246 | 11 Q61435 | Q61435 mus musculus |
| 3 | 1221.5 | 91.2 | 1669 | 11 Q9QZS0 | Q9qzs0 mus musculus |
| 4 | 1164 | 86.9 | 230 | 11 Q63122 | Q63122 rattus norv |
| 5 | 1158 | 86.4 | 212 | 6 Q28512 | Q28512 macaca mula |
| 6 | 1103 | 82.3 | 212 | 6 Q28567 | Q28567 ovis aries |
| 7 | 1084 | 80.9 | 210 | 6 Q28273 | Q28273 canis fami |
| 8 | 1072 | 80.0 | 203 | 6 Q28682 | Q28682 oryctolagus |
| 9 | 1055 | 78.7 | 203 | 6 Q29032 | Q29032 sus scrofa |
| 10 | 960 | 71.6 | 1075 | 4 Q86X41 | Q86x41 homo sapien |
| 11 | 960 | 71.6 | 1621 | 4 Q9H4R9 | Q9h4r9 homo sapien |
| 12 | 957.5 | 71.5 | 979 | 13 Q919K3 | Q919k3 gallus gall |
| 13 | 952.5 | 71.1 | 253 | 11 Q61436 | Q61436 mus musculus |
| 14 | 952.5 | 71.1 | 585 | 11 Q80V57 | Q80v57 mus musculus |
| 15 | 952.5 | 71.1 | 799 | 11 Q8BNS7 | Q8bns7 mus musculus |
| 16 | 952.5 | 71.1 | 1691 | 11 Q9ESQ2 | Q9esq2 mus musculus |

Q9nub7 homo sapien
Q9nyc5 homo sapien
Q9n1q8 mus musculu
Q8n188 homo sapien
Q8hyc1 canis fami
Q86622 canis fami
Q28271 canis fami
Q07265 strongyloce
Q61430 mus musculu
Q17163 brugia mala
P70165 mus musculu
Q26640 strongyloce
Q14052 homo sapien
Q91vi3 mus musculu
Q9esq1 mus musculu
Q9n97 mus musculu
Q9vmv4 drosophila
Q9n99 mus musculu
Q64457 mus musculu
Q8t754 anopheles g
Q9gq1 hydra atten
Q28272 canis fami
Q28274 canis fami
Q29468 canis fami
Q9gv24 sarcophaga
Q18407 drosophila
Q9vmv5 drosophila
O09238 pseudocorti
Q8ccdc6 mus musculu

ALIGNMENTS

RESULT 1

Q9NYC4 PRELIMINARY; PRT; 245 AA.

ID Q9NYC4

AC Q9NYC4; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Tumstatin (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Maeshina Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,

RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.,

RT 'Distinct anti-tumor properties of a type IV collagen domain derived

RT from basement membrane.';

RL J. Biol. Chem. 0:0-0(2000).

DR EMBL; AF258351; AAF72632.1; "

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0003676; P:nucleic acid binding; IEA.

DR InterPro; IPR001442; ProcollagN_C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagN_C4; 1.

DR SMART; SM00111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

FT NON_TER

SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 100.0%; Score 1340; DB 4; Length 245;

Best Local Similarity 100.0%; Pred. No. 2.3e-129;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPECTVPLYSFSLFVQGNORAGQD 60

DB 2 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPECTVPLYSFSLFVQGNORAGQD 61

```

QY 61 LGTLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
DB 62 LGTLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 121
QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSCL 180
DB 122 RCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSCL 181
QY 181 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 240
DB 182 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 241
QY 241 MKKRH 244
DB 242 MKKRH 245

RESULT 2
ID Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: Sequence, distribution, association with laminins, and
RT developmental switches.";
RL J. Cell Biol. 127:879-891(1994).
EN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; Z35166; CAA84529.1; -.
DR PIR; I48302;
DR MGI; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
FT NON_TER
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 91.2%; Score 1221.5; DB 11; Length 246;
Best Local Similarity 90.6%; Pred. No. 3.4e-117;
Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

QY 1 GLKGRGDSGPATWT-TRGFVFRHSQTTAIPSCPGTVELYSGFSFLFVQGNRAHQ 59
DB 2 GLKGNPGRGTATGTRVRGFIIFTHSQTTAIPSCPGTQPLYSGFSLLFVQGNKRAHQ 61
QY 60 DLGTGLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYI 119
DB 62 DLGTGLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYI 121
QY 120 SRTCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSCL 179
DB 122 SRTCTVCEGPAIAIAVHSQTTAIPCPQDWVSLWKGRFSIMFTSAGSEGTGOALASPGSCL 181
QY 180 DLGTGLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYI 119
DB 182 DLGTGLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYI 121
QY 180 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 239
DB 182 SRTCTVCEGPAIAIAVHSQTTAIPCPQDWVSLWKGRFSIMFTSAGSEGTGOALASPGSCL 179
DB 122 SRTCTVCEGPAIAIAVHSQTTAIPCPQDWVSLWKGRFSIMFTSAGSEGTGOALASPGSCL 181

```

```

QY 180 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 239
DB 182 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 241
QY 240 MKKRH 244
DB 242 MKKRH 246

RESULT 3
ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.
AC Q9QZS0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Alpha 3 collagen IV.
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.P.B., Miner J.H., Overbeek P.A., Weisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR PIR; I48302; I48302.
DR EMBL; AF169387; AAD50449.1; -.
DR MGI; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
KW Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 91.2%; Score 1221.5; DB 11; Length 1669;
Best Local Similarity 90.6%; Pred. No. 3.3e-116;
Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

QY 1 GLKGRGDSGPATWT-TRGFVFRHSQTTAIPSCPGTVELYSGFSFLFVQGNRAHQ 59
DB 1425 GLKGNPGRGTATGTRVRGFIIFTHSQTTAIPSCPGTQPLYSGFSLLFVQGNKRAHQ 1484
QY 60 DLGTGLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYI 119
DB 1485 DLGTGLGSLQRFTHMPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYI 1544
QY 120 SRTCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSCL 179
DB 1545 SRTCTVCEGPAIAIAVHSQTTAIPCPQDWVSLWKGRFSIMFTSAGSEGTGOALASPGSCL 1604
QY 180 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 239
DB 1605 EFRASPFLCHGRGTCNYNSYFWLASLNPRMFRKPIPSYKAGLEKIISRCQVCM 1664
QY 240 MKKRH 244
DB 1665 MKKRH 1669

```



```
DR InterPro; IPR001442; Procollagen4 C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 212 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 82.3%; Score 1103; DB 6; Length 212;
Best Local Similarity 92.4%; Pred. No. 4.3e-105;
Matches 195; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 33 SCPEGTVPVLYSGFSLFVQGNQAHGQDLGTLGSCLORETTPFLFCNVNDVCFASRND 92
DB 1 SCPEGTVPVLYSGFSLFVQGNQAHGQDLGTLGSCLORETTPFLFCNVNDVCFASRND 60

QY 93 YSYWLSTPALMPNMMAPIITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 152
DB 61 YSYWLSTPALMPNMMAPIITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 120

QY 153 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 212
DB 121 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 180

QY 213 ERMFKEPIPTVKAGELEKIISRCQVCMKKR 243
DB 181 QRMFKPIPTVKAGELEKIISRCQVCMKKR 211

RESULT 7
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
of collagen type IV. Evidence from a canine model of X-linked
nephritis with a COL4A3 gene mutation."
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagen4 C.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 210 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA82633D CRC64;

Query Match 80.9%; Score 1084; DB 6; Length 210;
Best Local Similarity 91.9%; Pred. No. 3.8e-103;
Matches 193; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 23 TRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQAHGQDLGTLGSCLORETTPFLFCNVN 82
DB 1 TRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQAHGQDLGTLGSCLORETTPFLFCNVN 60

QY 83 DVCNFEASRNDYSYWLSTPALMPNMMAPIITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIP 142
DB 61 NVNCFASRNDYSYWLSTPALMPNMMAPIITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIP 120

QY 143 PCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNYNS 202
DB 121 SCPNGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNYNS 180

QY 203 YSWFLASLNPFRMPKPIPTVKAGELEKI 232
DB 181 YSWFLASLNPFRMPKPIPTVKAGELEKI 210

RESULT 8
Q28682 PRELIMINARY; PRT; 203 AA.
AC Q28682;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RT "Properties and sequences of the Goodpasture antigen of different
mammals."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; I47283; AAA91893.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagen4 C.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 203 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 80.0%; Score 1072; DB 6; Length 203;
Best Local Similarity 95.1%; Pred. No. 6.2e-102;
Matches 193; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 33 SCPEGTVPVLYSGFSLFVQGNQAHGQDLGTLGSCLORETTPFLFCNVNDVCFASRND 92
DB 1 SCPEGTVPVLYSGFSLFVQGNQAHGQDLGTLGSCLORETTPFLFCNVNDVCFASRND 60

QY 93 YSYWLSTPALMPNMMAPIITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 152
DB 61 YSYWLSTPALMPNMMAPIITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 120

QY 153 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 212
DB 121 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 180

QY 213 ERMFKPIPTVKAGELEKIISR 235
DB 181 QRMFKPIPTVKAGELEKIISR 203
```

```
RESULT 9
Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OC NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
EMBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4 C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
DR Collagen.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
FT NON_TER 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 78.7%; Score 1055; DB 6; Length 203;
Best Local Similarity 93.1%; Pred. No. 3.5e-100;
Matches 189; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 33 SCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLGSLQRTTTPFLFCNVNDVCFASRND 92
DB 1 SCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLGSLQRTTTPFLFCNVNDVCFASRND 60

QY 93 YSYWLSTPALMPMNAITGRALBPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 152
DB 61 YSYWLSTPALMPMNAITGRALBPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 120

QY 153 KGFSFIMFTSAGSGTGQALASPGSCLEEFRAFPFLECHGRGTCNVYNSYSFWLASLNP 212
DB 121 KGFSFIMFTSAGSGTGQALASPGSCLEEFRAFPFLECHGRGTCNVYNSYSFWLASLNP 180

QY 213 ERMRKPIPTVKAGELEKIISR 235
DB 181 ERMRKPIPTVKAGELEKIISR 203

RESULT 10
Q86X41 PRELIMINARY; PRT; 1075 AA.
AC Q86X41;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
```

```
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC047305; AA47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
DR Collagen.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BDS52503D CRC64;

Query Match 71.6%; Score 960; DB 4; Length 1075;
Best Local Similarity 69.0%; Pred. No. 1.4e-89;
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPECTVLYSGFSLFVQGNQRAHGQD 60
DB 835 GLPGSMGPPGTPS--VDHGFIVTRHSQTTIDDPQCFSGTKILYHGYSLLYVQGNRAHGQD 892

QY 61 LGTGLSCLQRTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNAITGRALBPYIS 120
DB 893 LGTAGSCLRKFTSTMPFLFCNVNNVCFASRNDYSYWLSTPEPMPMSAPITGENIRPFIS 952

QY 121 RCTCEGPAIAIAVHSQTTDIPPCPHGWISLWKFIMFTSAGSGTGQALASPGSCLE 180
DB 953 RCACEAPAMVAVHSQTIQIPPCSGWSSLMICYFSFVMTSAGSGTGQALASPGSCLE 1012

QY 181 EFRASPFLECHGRGTCNVYNSYSFWLASLNPFRKPIPTVKAGELEKIISRQVCM 240
DB 1013 EFRASPFLECHGRGTCNVYANAYFWLATIERSEMFKKPTSTLKAGELRTHVSRQVCM 1072

QY 241 KK 242
DB 1073 RR 1074

RESULT 11
Q9H4R9 PRELIMINARY; PRT; 1621 AA.
AC Q9H4R9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE BA472K17.2 (Collagen type IV alpha 1 (fragment)).
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AL390755; CAC13153.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g_helix; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Collagen.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1621 AA; 155705 MW; 73F6F901CDOEDBA2 CRC64;
```

```

Query Match      71.6%; Score 960; DB 4; Length 1621;
Best Local Similarity 69.0%; Pred. No. 2.3e-89;
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKKGKDGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLVQGNQRAHQD 60
DB 1381 GLPGSMGPPGTPS--VDHGLVTRHSQTTDDPCPSGKILYHGYSLLVQGNRAHQD 1438
QY 61 LGTLGSCLOQFTTNPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMMAPIITGRALEPYIS 120
DB 1439 LGTAGSCLRFSTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMMAPIITGRIRPIS 1498
QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGSFIMFTSAGSEGTGOALASPGSCLE 180
DB 1499 RCACVCEAPAMVAHSQTTQIIPCPSGWSSLMIGYSFVNMHTSAGAGSGQALASPGSCLE 1558
QY 181 EFRASPLECHGRGTCNYSNSYSFWLASINPERMPKPIPTSVKAGELEKIIISRCQVCM 240
DB 1559 EFRASPFIECHGRGTCNYSNSYSFWLASINPERMPKPIPTSVKAGELEKIIISRCQVCM 1618
QY 241 KK 242
DB 1619 RR 1620

RESULT 12
Q919K3 PRELIMINARY; PRT; 979 AA.
AC Q919K3;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal basal lamina.";
RL Dev. Biol. 0:0-0(2000).
DR EMBL; AF239838; AAR44681.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg.Helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01381; Collagen; 12.
DR ProDom; PD000007; Clg.helix; 2.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER
SQ SEQUENCE 1 1
1 5020 MW; 5B1017D911ED4299 CRC64;

Query Match      71.5%; Score 957.5; DB 13; Length 979;
Best Local Similarity 67.3%; Pred. No. 2.3e-89;
Matches 167; Conservative 33; Mismatches 41; Indels 7; Gaps 1;

QY 1 GLKKGKDGSP-----ATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLVQGN 53
DB 730 GIPGPPGEGLPAGMGPGASVAHGFLVTRHSQTTTEEPQCPGTLIHGYSLLVQGN 789
QY 54 QRAHQDLGLGSCLOQFTTNPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMMAPIITGR 113
DB 790 ERAHQDLGTAGSCLRFSTMPFLFCNVNDVCFNFSRNDYSYWLSTPEPMPMSMAPIITGB 849
QY 114 ALEYSIRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGSFIMFTSAGSEGTGOALA 173
DB 114 ALEYSIRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGSFIMFTSAGSEGTGOALA 173

```

```

Db 850 SIRPFISRCVCEAPAMVAIAVHSQTTQIIPCPPEGWSSLMIGYSFVNMHTSAGAGSGQALA 909
QY 174 SPGSCLEERASPLECHGRGTCNYSNSYSFWLASINPERMPKPIPTSVKAGELEKII 233
DB 910 SPGSCLEERASPFIECHGRGTCNYSNSYSFWLASINPERMPKPIPTSVKAGELEKII 969
QY 234 SRCQVCMK 241
DB 970 SRCQVCMR 977

RESULT 13
Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: Sequence, distribution, association with laminins, and developmental switches.";
RL J. Cell Biol. 127:879-891(1994).
DR EMBL; Z35168; CAA84531.1; -.
DR F1R; I48304; I48304.
DR MGP; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match      71.1%; Score 952.5; DB 11; Length 253;
Best Local Similarity 68.7%; Pred. No. 1.5e-89;
Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY 1 GLKKGKDGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLVQGNQRAHQ 59
DB 10 GPDGLQGPDPGPGTTSVAHGFLVTRHSQTTTEAPCPRGTVHIYEGSLLVQGNKRAHQ 69
QY 60 DLGTLGSCLOQFTTNPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMMAPIITGRALEPYI 119
DB 70 DLGTAGSCLRFSTMPFLFCNVNDVCFNFSRNDYSYWLSTPEPMPNMNEFLKQGSIQPFI 129
QY 120 SRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGSFIMFTSAGSEGTGOALASPGSCL 179
DB 130 SRCVCEAPAMVAIAVHSQTTQIIPCPQGNDSLMIGYSFVNMHTSAGAGSGQALASPGSCL 189
QY 180 EFRASPLECHGRGTCNYSNSYSFWLASINPERMPKPIPTSVKAGELEKIIISRCQVC 239
DB 190 EFRASPFIECHGRGTCNYSNSYSFWLASINPERMPKPIPTSVKAGELEKIIISRCQVC 249
QY 240 MKK 242
DB 250 MKR 252

RESULT 14
Q80V57 PRELIMINARY; PRT; 585 AA.
ID Q80V57

```



```
AC Q80V57;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Col4a5 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FVB/N; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan M., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uadin T.B., Tothiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC043317; AAH43317.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagenC4; 2.
DR SMART; SM00111; C4; 2.
SQ SEQUENCE 585 AA; 58283 MW; 26774FE364E7FD8D CRC64;

Query Match 71.1%; Score 952.5; DB 11; Length 585;
Best Local Similarity 68.7%; Pred. No. 4.1e-89;
Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY 1 GLKKGKGDSPATWT-TRGFVTRHSQTTAIPSCPGTVPYSGFSLFVQGNORAHGQ 59
Db 342 GPDLQGPDPGGTTSVAHGFLLITRHSQTTAIPSCPGTVPYSGFSLFVQGNORAHGQ 401
QY 60 DLGTGLSCLQRFTTTPFLFCNVNDVNCNFASRNDYSYWLSTPMPMNMMAPIITGRALEPYI 119
Db 402 DLGTAGSCLRRFSTMPFNCNNVNCNFASRNDYSYWLSTPMPMNMMAPIITGRALEPYI 461
QY 120 SRCVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGSFIMFTSAGSEGTQALASPGSCL 179
Db 462 SRCVCECAPAVIAVHSQTTQIPHCPCQGDLSLWIGYSFMWHTSAGAEQSGQALASPGSCL 521
QY 180 EEFRAFPLECHGRGTCNYYNSYSYFWLASINPERMFRKPIPTVKGAELEKTIISRCQVC 239
Db 522 EEFRAFPFISCHGRGTCNYYANSYSYFWLAIIVDSMDNFNKPSQETLKGADLRTRISRCQVC 581
QY 240 MKK 242
Db 582 MKR 584
```

```
RESULT 15
Q8BNS7
ID Q8BNS7 PRELIMINARY; PRT; 799 AA.
AC Q8BNS7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGI; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 71.1%; Score 952.5; DB 11; Length 799;
Best Local Similarity 68.7%; Pred. No. 5.9e-89;
Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY 1 GLKKGKGDSPATWT-TRGFVTRHSQTTAIPSCPGTVPYSGFSLFVQGNORAHGQ 59
Db 556 GPDLQGPDPGGTTSVAHGFLLITRHSQTTAIPSCPGTVPYSGFSLFVQGNORAHGQ 615
QY 60 DLGTGLSCLQRFTTTPFLFCNVNDVNCNFASRNDYSYWLSTPMPMNMMAPIITGRALEPYI 119
Db 616 DLGTAGSCLRRFSTMPFNCNNVNCNFASRNDYSYWLSTPMPMNMMAPIITGRALEPYI 675
QY 120 SRCVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGSFIMFTSAGSEGTQALASPGSCL 179
Db 676 SRCVCECAPAVIAVHSQTTQIPHCPCQGDLSLWIGYSFMWHTSAGAEQSGQALASPGSCL 735
QY 180 EEFRAFPLECHGRGTCNYYNSYSYFWLASINPERMFRKPIPTVKGAELEKTIISRCQVC 239
Db 736 EEFRAFPFIECHGRGTCNYYANSYSYFWLAIIVDSMDNFNKPSQETLKGADLRTRISRCQVC 795
QY 240 MKK 242
Db 796 MKR 798
```

Search completed: April 5, 2004, 07:03:57
Job time : 149.518 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 219.777 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 1340

Sequence: 1 GLKKGKDGSGSPATWTRGF.....KAGELEKIISRQVCVKRKH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 1340 | 100.0 | 244 | 5 | ABG79219 Human Goo |
| 2 | 1340 | 100.0 | 244 | 5 | Aau75595 Human typ |
| 3 | 1340 | 100.0 | 244 | 6 | ADA20225 Human typ |
| 4 | 1340 | 100.0 | 245 | 3 | AAY67942 Human typ |
| 5 | 1340 | 100.0 | 245 | 5 | Aau75589 Human typ |
| 6 | 1340 | 100.0 | 1670 | 7 | ADD47063 Human Pro |
| 7 | 1337 | 99.8 | 244 | 5 | ABG79218 Human typ |
| 8 | 1336 | 99.7 | 244 | 5 | ABG79217 Human typ |
| 9 | 1317 | 98.3 | 254 | 5 | Aau75598 Human typ |
| 10 | 1314 | 98.1 | 268 | 2 | AAY31993 Type IV c |
| 11 | 1314 | 98.1 | 268 | 3 | AAY97555 Human alp |
| 12 | 1271 | 94.9 | 232 | 7 | ADC17697 Human typ |
| 13 | 1210.5 | 90.3 | 471 | 2 | AAY44171 Bovine ty |
| 14 | 1210.5 | 90.3 | 471 | 3 | AAY56783 Bovine al |
| 15 | 1210.5 | 90.3 | 471 | 4 | AAY56783 Bovine al |
| 16 | 1199.5 | 89.5 | 471 | 2 | AAY93163 Partial s |
| 17 | 1192 | 89.0 | 218 | 2 | AAY44172 Human typ |
| 18 | 1192 | 89.0 | 218 | 3 | AAY56784 Human alp |
| 19 | 1192 | 89.0 | 218 | 4 | AAY56784 Human alp |
| 20 | 1170 | 87.3 | 218 | 2 | AAY93164 Partial s |
| 21 | 1164 | 86.9 | 230 | 7 | ADD47061 Rat Prote |
| 22 | 1052 | 78.5 | 191 | 6 | ADA20260 Human tum |
| 23 | 1050 | 78.4 | 191 | 5 | Aau75596 Human typ |
| 24 | 960 | 71.6 | 406 | 3 | AAB58169 Lung canc |
| 25 | 960 | 71.6 | 1669 | 4 | AAM40863 Human pol |

| | | | | | |
|----|-------|------|------|---|--------------------|
| 26 | 960 | 71.6 | 1669 | 5 | ABB90760 Human Tum |
| 27 | 960 | 71.6 | 1669 | 6 | ABU54467 Human tum |
| 28 | 960 | 71.6 | 1672 | 4 | AM35077 Human pol |
| 29 | 951 | 71.0 | 1669 | 5 | ABBS7334 Mouse isc |
| 30 | 943.5 | 70.4 | 772 | 2 | AAR23873 Human alp |
| 31 | 943.5 | 70.4 | 772 | 2 | AAM09643 Human typ |
| 32 | 943.5 | 70.4 | 1685 | 4 | ABG04839 Novel hum |
| 33 | 943.5 | 70.4 | 1693 | 4 | ABG15619 Novel hum |
| 34 | 939 | 70.1 | 229 | 3 | AAY67943 Human typ |
| 35 | 939 | 70.1 | 229 | 3 | AAY75587 Human typ |
| 36 | 939 | 70.1 | 229 | 6 | ADA20217 Human typ |
| 37 | 939 | 70.1 | 229 | 7 | ADC17695 Human typ |
| 38 | 939 | 70.1 | 260 | 2 | AAY31991 Type IV c |
| 39 | 939 | 70.1 | 260 | 3 | AAY97553 Human alp |
| 40 | 938.5 | 70.0 | 264 | 2 | AAY31995 Type IV c |
| 41 | 938.5 | 70.0 | 264 | 3 | AAY97557 Human alp |
| 42 | 937.5 | 70.0 | 309 | 3 | AAB54044 Human pan |
| 43 | 928 | 69.3 | 229 | 7 | ADC17699 Human typ |
| 44 | 923 | 68.9 | 229 | 1 | AAP93524 Complete |
| 45 | 880 | 65.7 | 211 | 3 | AAY95918 Human Goo |

ALIGNMENTS

RESULT 1

ABG79219

ID ABG79219 standard; protein; 244 AA.

XX AC ABG79219;

XX DT 15-NOV-2002 (first entry)

XX DE Human Goodpasture disease-related protein.

XX KW Goodpasture antigen binding protein; Goodpasture syndrome;

XX KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;

XX KW autoimmunity condition; phosphorylation; myelin basic protein; MBP;

XX KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;

XX KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;

XX KW pemphigoid; lichen planus.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200261430-A2.

XX PD 08-AUG-2002.

XX PF 31-JAN-2002; 2002WO-EP001010.

XX PR 31-JAN-2001; 2001US-0265249P.

XX PA (SAUS/) SAUS J.

XX PI Saus J;

XX DR WPI; 2002-619280/66.

XX XX Identifying candidate compounds for treating autoimmune conditions, e.g.

XX PT Goodpasture syndrome or lupus, comprises identifying compounds that

XX PT reduce phosphorylation of, or formation of conformational isomers of,

XX PT target proteins.

XX PS Disclosure; Page 213-214; 217pp; English.

XX CC The invention relates to identifying candidate compounds to treat an

XX CC autoimmune condition by identifying compounds that reduce phosphorylation

XX CC of a first target protein (I) (which is selected from Goodpasture antigen

XX CC binding protein (GPP), an alpha3 type IV collagen non-collagenous (NCI)

XX CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-

XX CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising

XX CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of

conformational isomers of the second target protein (II) (selected from an alpha3 type IV collagen NCI domain polypeptide and myelin basic protein, MBP). Also included are (1) an isolated type IV collagen alpha3 NCI domain conformational isomer, which has an amino acid sequence identical to the wild type alpha3 type IV collagen NCI domain, is stabilised by disulphide bonds, and has a molecular weight in a non-reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kDa, and in a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated type IV collagen alpha3 NCI domain. The human gene for GPAP is located on chromosome 5q13. The method is useful for treating autoimmune conditions, such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present CC sequence represents a Goodpasture syndrome related protein or peptide XX
SQ Sequence 244 AA;

Query Match 100.0%; Score 1340; DB 5; Length 244;
Best Local Similarity 100.0%; Pred. No. 2.9e-133;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPECTVPLYSFGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPECTVPLYSFGFSLFVQGNQRAHQD 60
QY 61 LGTLGSLCLORETTMPFLFCNVNDVNCNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
DB 61 LGTLGSLCLORETTMPFLFCNVNDVNCNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
QY 121 RCTVCEGPAIAIAVHSQTTPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
DB 121 RCTVCEGPAIAIAVHSQTTPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
QY 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
DB 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
QY 241 KKRH 244
DB 241 KKRH 244

RESULT 2
AAU75595
ID AAU75595 standard; protein; 244 AA.
AC AAU75595;

XX 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tumstatin 334.

KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000365.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
PT treating disorders involving angiogenesis.

XX Example 33; Page; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha4, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments inducing angiogenesis or cell proliferation in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 3 chain mutant, Tumstatin 334, which consists of residues 2-245 of Tumstatin. Note: The present sequence is not shown in the specification but is derived from the wild type human Tumstatin sequence given in figure 18A (see AAU75589)

XX SQ Sequence 244 AA;

Query Match 100.0%; Score 1340; DB 5; Length 244;
Best Local Similarity 100.0%; Pred. No. 2.9e-133;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPECTVPLYSFGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPECTVPLYSFGFSLFVQGNQRAHQD 60
QY 61 LGTLGSLCLORETTMPFLFCNVNDVNCNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
DB 61 LGTLGSLCLORETTMPFLFCNVNDVNCNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
QY 121 RCTVCEGPAIAIAVHSQTTPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
DB 121 RCTVCEGPAIAIAVHSQTTPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
QY 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
DB 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
QY 241 KKRH 244
DB 241 KKRH 244

RESULT 3
ADA20225
ID ADA20225 standard; protein; 244 AA.
XX AC ADA20225;
XX AC ADA20225;
XX DT 20-NOV-2003 (first entry)
XX DE Human type IV collagen alpha 3 chain partial protein sequence.
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 1..244
XX FT Region /note= "Tumstatin"
XX FT Region 1..124
XX FT Region /note= "Tumstatin 333; pET22b-alpha 3 (IV) NCI region"
XX FT Region 1..19
XX FT Region /note= "T1 peptide"
XX FT Region 28..42
XX FT Region /note= "First Goodpasture epitope"
XX FT Region 53..72
XX FT Region /note= "T2 peptide"
XX FT Region 68..88
XX FT Region /note= "T3 peptide"
XX FT Region 83..102
XX FT Region /note= "T4 peptide"
XX FT Region 98..116
XX FT Region /note= "T5 peptide"
XX FT Region 113..132
XX FT Region /note= "T6 peptide"
XX FT Region 125..244
XX FT Region /note= "Tumstatin 334"
XX FT Region 139..152
XX FT Region /note= "Second Goodpasture epitope"
XX PN WO2003059257-A2.
XX PD 24-JUL-2003.
XX PF 20-DEC-2002; 2002WO-US040398.
XX PR 21-DEC-2001; 2001US-00032221.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2003-587256/55.
XX DR N-PSDB; ADA20224.
XX PT New peptide, useful for preparing a composition for inhibiting tumor
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX PS Claim 52; Fig 18; 240pp; English.
XX CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumor growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is the partial amino acid sequence of the alpha 3 chain of human
CC type IV collagen. The "tumstatin" protein of the invention was derived
CC from this protein and comprises the full length of the present sequence.
XX SQ Sequence 244 AA;
XX Query Match 100.0%; Score 1340; DB 6; Length 244;
XX Best Local Similarity 100.0%; Pred No. 2.9e-133;
XX Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPILYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPILYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSCLORETTMPFLFCNVNDYCNFASRNDYSYWLSTPALMPMNAPITGRALEPVIS 120
DB 61 LGTLGSCLORETTMPFLFCNVNDYCNFASRNDYSYWLSTPALMPMNAPITGRALEPVIS 120
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALASPGSCLE 180
DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALASPGSCLE 180
QY 181 EFRASPFLCHGRGTCNVYNSYSYFWLASLNPERMFRKPISTVKAGELEKIIIRCOQVM 240
DB 181 EFRASPFLCHGRGTCNVYNSYSYFWLASLNPERMFRKPISTVKAGELEKIIIRCOQVM 240
QY 241 KKRH 244
DB 241 KKH 244
RESULT 4
AAV67942
ID AAV67942 standard; protein; 245 AA.
XX AC AAV67942;
XX DT 03-APR-2000 (first entry)
XX DE Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.
XX KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
KW benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
KW ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;
KW myocardial angiogenesis; plaque neovascularisation; angiofibroma;
KW atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
KW contraception; obesity.
XX OS Homo sapiens.
XX PN WO9965940-A1.
XX PD 23-DEC-1999.
XX PF 17-JUN-1999; 99WO-US013737.
XX PR 17-JUN-1998; 98US-0089689P.
XX PR 25-MAR-1999; 99US-0126175P.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2000-097708/08.
XX DR N-PSDB; AA257158.
XX PT Anti-angiogenic proteins comprising the NCI domain of the alpha 1, 2 or 3

PT chain of Type IV collagen used in, e.g. treatment of benign tumors and
 XX rheumatoid arthritis.
 XX
 XX Claim 32; Fig 16B; 117pp; English.
 XX
 CC The present sequence represents the human type IV collagen alpha 3 chain.
 CC The present invention describes an isolated protein chosen from the NCI
 CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a
 CC fragment, analogue, derivative or mutant, which has anti-angiogenic
 CC properties. The anti-angiogenic proteins, multimers and chimeras are
 CC useful for inhibiting angiogenic activity in mammalian tissue, especially
 CC for treating diseases chosen from angiogenesis-dependent cancers, benign
 CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular
 CC angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints, atherosclerosis,
 CC angiofibroma, wound granulation, intestinal adhesions, arterioleclerosis,
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori
 CC ulcers, dialysis graft vascular access stenosis, contraception and
 CC obesity. The compositions can be used to inhibit a disease characterised
 CC by angiogenic activity, in conjunction with radiation therapy,
 CC chemotherapy or immunotherapy
 XX
 SQ Sequence 245 AA;

Query Match 100.0%; Score 1340; DB 3; Length 245;
 Best Local Similarity 100.0%; Pred. No. 2.9e-133;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60
 Db 2 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 61
 QY 61 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYIS 120
 Db 62 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYIS 121
 QY 121 RCTVCEGPAIAVAHSQTTPDIPPCPHGWISLWKGFSPFIMFTSAGSEGTGQALASPGSCLE 180
 Db 122 RCTVCEGPAIAVAHSQTTPDIPPCPHGWISLWKGFSPFIMFTSAGSEGTGQALASPGSCLE 181
 QY 181 EFRASPFLECHGRGTCNVYNSYSFWLASLNPFRMPKPIPTVKAGELEKIIISRCQVCM 240
 Db 182 EFRASPFLECHGRGTCNVYNSYSFWLASLNPFRMPKPIPTVKAGELEKIIISRCQVCM 241
 QY 241 KKRH 244
 Db 242 KKRH 245

RESULT 5
 AAU75589
 ID AAU75589 standard; protein; 245 AA.

XX AAU75589;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain, 'Tumstatin'.
 XX
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200151523-A2.
 XX
 PD 19-JUL-2001.
 XX
 PP 08-JAN-2001; 2001WO-US0000565.
 XX
 PR 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX
 PI Kalluri R;
 XX
 DR WPI, 2002-188037/24.
 DR N-FSDB; ABK15365.
 XX
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.
 PT
 PT Claim 29; Fig 18B; 205pp; English.
 PS
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain
 XX
 SQ Sequence 245 AA;

Query Match 100.0%; Score 1340; DB 5; Length 245;
 Best Local Similarity 100.0%; Pred. No. 2.9e-133;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60
 Db 2 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 61
 QY 61 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYIS 120
 Db 62 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYIS 121
 QY 121 RCTVCEGPAIAVAHSQTTPDIPPCPHGWISLWKGFSPFIMFTSAGSEGTGQALASPGSCLE 180
 Db 122 RCTVCEGPAIAVAHSQTTPDIPPCPHGWISLWKGFSPFIMFTSAGSEGTGQALASPGSCLE 181
 QY 181 EFRASPFLECHGRGTCNVYNSYSFWLASLNPFRMPKPIPTVKAGELEKIIISRCQVCM 240

Db 182 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMRKPISTVKAGELEKIISRCQVCM 241
QY 241 KKRH 244
Db 242 KKRH 245

RESULT 6
ADD47063
ID ADD47063 standard; protein; 1670 AA.
XX
AC ADD47063;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein NP_000082, SEQ ID NO 12751.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
FN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GENO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; NP_000082.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pt_sequences.

XX SQ Sequence 1670 AA;
Query Match 100.0%; Score 1340; DB 7; Length 1670;
Best Local Similarity 100.0%; Pred. No. 3.7e-132;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTTTRGFVTRHSQTTAIPSCPGTVPFLYSGFSLFVQGNQRAHQD 60
Db 1427 GLKGRGDSGSPATWTTTRGFVTRHSQTTAIPSCPGTVPFLYSGFSLFVQGNQRAHQD 1486
QY 61 LGTLGSCLOQRTTWPFLFCNVNDVCFASRNDYSYWLSTPALMPMMAPIITGRALBPYIS 120
Db 1487 LGTLGSCLOQRTTWPFLFCNVNDVCFASRNDYSYWLSTPALMPMMAPIITGRALBPYIS 1546
QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWSILWKGFIFIMFTSAGSEGTGQALASPGSCLE 180
Db 1547 RCTVCEGPAIAIAVHSQTTDIPCPHGWSILWKGFIFIMFTSAGSEGTGQALASPGSCLE 1606
QY 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMRKPISTVKAGELEKIISRCQVCM 240
Db 1607 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMRKPISTVKAGELEKIISRCQVCM 1666
QY 241 KKRH 244
Db 1667 KKRH 1670

RESULT 7
ABG79218
ID ABG79218 standard; protein; 244 AA.
XX
AC ABG79218;
XX
DT 15-NOV-2002 (first entry)
XX
DE Human type IV collagen NC1 domain mutant, alpha3(IV)NC1Aala9.
XX
KW Goodpasture antigen binding protein; Goodpasture syndrome;
KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
KW alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;
KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
KW pemphigoid; lichen planus; human; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FN WO200261430-A2.
XX
PD 08-AUG-2002.
XX
PF 31-JAN-2002; 2002WO-EP001010.
XX
PR 31-JAN-2001; 2001US-0265249P.
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2002-619280/66.
DR N-PSDB; ABS64503.
XX
PT Identifying candidate compounds for treating autoimmune conditions, e.g.
PT Goodpasture syndrome or lupus, comprises identifying compounds that
PT reduce phosphorylation of, or formation of conformational isomers of,
PT target proteins.
XX
PS Claim 21; Page 212-213; 217pp; English.
XX
CC The invention relates to identifying candidate compounds to treat an
CC autoimmune condition by identifying compounds that reduce phosphorylation
CC of a first target protein (I) (which is selected from Goodpasture antigen

CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCl)
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-
 CC Ala-Thr-Trip-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
 CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly, or reduce formation of
 CC conformational isomers of the second target protein (II) (selected from
 CC an alpha3 type IV collagen NCl domain polypeptide and myelin basic
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
 CC NCl domain conformational isomer, which has an amino acid sequence
 CC identical to the wild type alpha3 type IV collagen NCl domain, is
 CC stabilised by disulphide bonds, and has a molecular weight in a non-
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
 CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
 CC type IV collagen alpha3 NCl domain. The human gene for GPBP is located on
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCl)
 CC domain (also known as the GP antigen) mutant
 XX
 SQ Sequence 244 AA;

Query Match 99.8%; Score 1337; DB 5; Length 244;
 Best Local Similarity 99.6%; Pred. No. 6.1e-133;
 Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKGRGDSGPATWTRGTFVTRHSQTTPAIPSCPEGTPLVSGSFVQGNQRAHQD 60
 DB 1 GLKGRGDSGPATWTRGTFVTRHSQTTPAIPSCPEGTPLVSGSFVQGNQRAHQD 60
 QY 61 LGTIGSCLOQRTTTPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
 DB 61 LGTIGSCLOQRTTTPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
 QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLE 180
 DB 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLE 180
 QY 181 EFRASPLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
 DB 181 EFRASPLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
 QY 241 KKRH 244
 DB 241 KKRH 244

RESULT 8
 ABG79217 standard; protein; 244 AA.
 XX
 AC ABG79217;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Human type IV collagen NCl domain mutant, alpha3 (IV)NCIAsp9.
 XX
 KW Goodpasture antigen binding protein; Goodpasture syndrome;
 KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
 KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
 KW alpha3 type IV collagen non-collagenous domain; NCl; multiple sclerosis;
 KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
 KW pemphigoid; lichen planus; human; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200261430-A2.
 XX
 PD 08-AUG-2002.
 XX
 PF 31-JAN-2002; 2002WO-EP001010.
 XX
 PR 31-JAN-2001; 2001US-0265249P.

XX (SAUS/) SAUS J.
 PA Saus J;
 XX
 PI Saus J;
 XX
 DR WPI: 2002-619280/66.
 DR N-PSDB; ABS64502.
 XX
 PT Identifying candidate compounds for treating autoimmune conditions, e.g.
 PT Goodpasture syndrome or lupus, comprises identifying compounds that
 PT reduce phosphorylation of, or formation of conformational isomers of,
 PT target proteins.
 XX
 PS Claim 21; Page 209-210; 217pp; English.
 XX
 CC The invention relates to identifying candidate compounds to treat an
 CC autoimmune condition by identifying compounds that reduce phosphorylation
 CC of a first target protein (I) (which is selected from Goodpasture antigen
 CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCl)
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-
 CC Ala-Thr-Trip-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
 CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
 CC conformational isomers of the second target protein (II) (selected from
 CC an alpha3 type IV collagen NCl domain polypeptide and myelin basic
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
 CC NCl domain conformational isomer, which has an amino acid sequence
 CC identical to the wild type alpha3 type IV collagen NCl domain, is
 CC stabilised by disulphide bonds, and has a molecular weight in a non-
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
 CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
 CC type IV collagen alpha3 NCl domain. The human gene for GPBP is located on
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCl)
 CC domain (also known as the GP antigen) mutant
 XX
 SQ Sequence 244 AA;

Query Match 99.7%; Score 1336; DB 5; Length 244;
 Best Local Similarity 99.6%; Pred. No. 7.8e-133;
 Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GLKGRGDSGPATWTRGTFVTRHSQTTPAIPSCPEGTPLVSGSFVQGNQRAHQD 60
 DB 1 GLKGRGDSGPATWTRGTFVTRHSQTTPAIPSCPEGTPLVSGSFVQGNQRAHQD 60
 QY 61 LGTIGSCLOQRTTTPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
 DB 61 LGTIGSCLOQRTTTPFLFCNVNDVCFNPNFASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
 QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLE 180
 DB 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLE 180
 QY 181 EFRASPLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
 DB 181 EFRASPLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240
 QY 241 KKRH 244
 DB 241 KKRH 244

RESULT 9
 AAU75598
 ID AAU75598 standard; protein; 254 AA.
 XX
 AC AAU75598;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain mutant, Tum-3.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.
 OS Homo sapiens.
 XX
 XX W0200151523-A2.
 XX
 PD 19-JUL-2001.
 XX
 XX 08-JAN-2001; 2001WO-US000565.
 XX
 XX 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA
 XX Kalluri R;
 PI
 XX
 XX WPI; 2002-188037/24.
 DR
 XX
 XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.
 PT
 XX
 XX Example 36; Page; 205pp; English.
 XX
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2, or
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-3, which consists of residues
 CC 133-244 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75599)
 XX
 XX Sequence 254 AA;

Query Match

98.3%; Score 1317; DB 5; Length 254;

Best Local Similarity 96.0%; Pred. No. 8.4e-131;
 Matches 243; Conservative 0; Mismatches 0; Indels 10; Gaps 1;
 QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTPAIPSCPECTVPLYSFGFSFLVQGNQRAHQD 60
 DB 1 GLKGRGDSGSPATWTRGFVTRHSQTTPAIPSCPECTVPLYSFGFSFLVQGNQRAHQD 61
 QY 61 -----LGTLSGCLQRTMPFLFCNVNDVCFASNDYSYWLSTALPMNMWAPI 110
 DB 62 LGTSGCLQRLGTLGSCLOQRTMPFLFCNVNDVCFASNDYSYWLSTALPMNMWAPI 121
 QY 111 TGRALEPYISRCTVCEGPAIAIAVHSQTTPAIPSCPECTVPLYSFGFSFLVQGNQRAHQD 170
 DB 122 TGRALEPYISRCTVCEGPAIAIAVHSQTTPAIPSCPECTVPLYSFGFSFLVQGNQRAHQD 181
 QY 171 ALASFGSCLEBFRAFPFLECHGRGTCNYNSNSYFWLASLNPMRFRKPIPTSVKAGELE 230
 DB 182 ALASFGSCLEBFRAFPFLECHGRGTCNYNSNSYFWLASLNPMRFRKPIPTSVKAGELE 241
 QY 231 KIISRCQVCKKR 243
 DB 242 KIISRCQVCKKR 254
 RESULT 10
 AAY31993
 ID AAY31993 standard; protein; 268 AA.
 XX
 AC AAY31993;
 XX
 DT 05-JAN-2000 (first entry)
 DE
 XX Type IV collagen NC1 domain alpha-3 monomer.
 KW Type IV collagen; NC1 domain; non-collagenous domain; human;
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
 KW rheumatoid arthritis; retinal neovascularization;
 KW choroidal neovascularization; macular degeneration;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidermal keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; soggrens; acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;
 KW sarcoid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu; AIDS;
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
 KW pemphigoid.
 OS Homo sapiens.
 XX Synthetic.
 Key Location/Qualifiers
 Peptide 1..17 /note= "EM40 signal peptide"
 Protein 18..268 /note= "mature protein"
 Peptide 18..25 /note= "affinity tag"
 Protein 26..268 /note= "NC1 alpha-3 monomer"
 WO9949885-A2.
 PN


```
XX PD 07-OCT-1999.
XX XX
XX PF 26-MAR-1999; 99WO-US006445.
XX XX
XX PR 27-MAR-1998; 98US-0079783P.
XX PR 29-OCT-1998; 98US-0106170P.
XX XX
XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX XX
XX PI Hudson BG, Sarraz MP;
XX XX
XX DR WPI: 1999-601297/51.
XX DR N-PSDB; AAZ20091.
XX XX
XX PT Inhibition of angiogenesis with non-collagenous alpha chain monomer
XX PT useful for treating e.g. tumor growth or metastasis, neovascularisation,
XX PT etc.
XX PS
XX PS Disclosure; Fig 17c; 56pp; English.
XX XX
XX CC This sequence represents a recombinant type IV collagen non-collagenous
XX CC (NC1) domain alpha-3 polypeptide composed of a B40 signal sequence
XX CC (which is cleaved from the mature protein) to facilitate protein
XX CC secretion, and a mature protein comprising an affinity tag (facilitates
XX CC purification and identification of the material) and the alpha-1 chain
XX CC monomer. The invention provides methods and kits for inhibiting
XX CC angiogenesis, tumour growth and metastasis, and endothelial cell
XX CC interaction with the extracellular matrix, each method comprising
XX CC contacting the tumour or animal tissue with 1 or more isolated type IV
XX CC collagen NC1 alpha chain monomer(s) selected from the group consisting of
XX CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-
XX CC 96). The monomers can be produced via recombinant protein expression. The
XX CC polynucleotides and polypeptides are used to treat an angiogenesis-
XX CC mediated disorder or condition, especially selected from solid and blood-
XX CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
XX CC neovascularization, choroidal neovascularization, macular degeneration,
XX CC corneal neovascularization, retinopathy of prematurity, corneal graft
XX CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
XX CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
XX CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
XX CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
XX CC simplex infections, herpes zoster infections, protozoan infections,
XX CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
XX CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's
XX CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,
XX CC sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,
XX CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
XX CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Behcets
XX CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic
XX CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
XX CC complications, abnormal proliferation of fibrovascular tissue,
XX CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
XX CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
XX CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)
XX XX
XX SQ Sequence 268 AA;

Query Match
Best Local Similarity 98.1%; Score 1314; DB 2; Length 268;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGSQSPATWTTGTFVTRHSQTATPSCPEGTVPLVYSGFSLFVCGNQRAGQDLGTL 64
DB 29 KRGSQSPATWTTGTFVTRHSQTATPSCPEGTVPLVYSGFSLFVCGNQRAGQDLGTL 88
QY 65 GSCLRFTTMTPLFCNVNDVNCNEARNDSYVWLSFTPLMPNMWAPITGRALEPIVSRCTV 124
DB 89 GSCLRFTTMTPLFCNVNDVNCNEARNDSYVWLSFTPLMPNMWAPITGRALEPIVSRCTV 148
QY 125 CEGPAIAIAVHSQTTDIPPCPHGWSLWKGSFIMFTSAGSEGTGQALASPGSCLEEFRA 184
|||||
```

Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGSQSPATWTRGVFTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTL 64
 DB 29 KRGSQSPATWTRGVFTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTL 88
 QY 65 GSCLORETTMPFLFCNVNDVNCNPNRNDYSYWLSTPALMPMNPATITGRALSPYISRCTV 124
 DB 89 GSCLORETTMPFLFCNVNDVNCNPNRNDYSYWLSTPALMPMNPATITGRALSPYISRCTV 148
 QY 125 CEGPAIAIVHSQTTDIPPCPHGWISLWKGFSTIMFTSAGSEGTGQALASPGSCLEEPPRA 184
 DB 149 CEGPAIAIVHSQTTDIPPCPHGWISLWKGFSTIMFTSAGSEGTGQALASPGSCLEEPPRA 208
 QY 185 SPFLECHGRGTCNYSNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCMKKRH 244
 DB 209 SPFLECHGRGTCNYSNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCMKKRH 268

RESULT 12
 ADC17697
 ID ADC17697 standard; protein; 232 AA.
 AC ADC17697;
 XX
 DT 18-DEC-2003 (first entry)
 DE Human type IV collagen alpha 3 chain protein SEQ ID NO:304.
 XX
 KW crystallised NCI domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO2003012122-A2.
 XX
 FD 13-FEB-2003.
 XX
 PF 26-JUL-2002; 2002WO-US023763.
 XX
 PR 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX
 PI Sundaramoorthy M, Hudson B;
 XX
 DR WPI; 2003-332730/31.
 XX
 PT New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 PS Disclosure; SEQ ID NO 304; 169pp; English.
 XX
 CC The present invention describes a crystallised NCI domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (5) inhibiting endothelial cell interaction with the

CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NCI
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
 CC antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents an amino acid sequence which is used in the exemplification of
 CC the present invention.

XX Sequence 232 AA;

Query Match 94.9%; Score 1271; DB 7; Length 232;
 Best Local Similarity 99.6%; Pred. No. 5.6e-126;
 Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 13 ATWTRGFVTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTLGSLQRT 72
 DB 1 ATWTRGFVTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTLGSLQRT 60
 QY 73 TMPFLFCNVNDVNCNPNRNDYSYWLSTPALMPMNPATITGRALSPYISRCTVCEGPAIAI 132
 DB 61 TMPFLFCNVNDVNCNPNRNDYSYWLSTPALMPMNPATITGRALSPYISRCTVCEGPAIAI 120
 QY 133 AVHSQTTDIPPCPHGWISLWKGFSTIMFTSAGSEGTGQALASPGSCLEEPPRAFPLECHG 192
 DB 121 AVHSQTTDIPPCPHGWISLWKGFSTIMFTSAGSEGTGQALASPGSCLEEPPRAFPLECHG 180
 QY 193 RGTGTCNYSNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCMKKRH 244
 DB 181 RGTGTCNYSNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCMKKRH 232

RESULT 13
 AAY44171
 ID AAY44171 standard; protein; 471 AA.
 XX
 AC AAY44171;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Bovine type IV collagen alpha3 chain protein.
 XX
 KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;
 KW Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.
 XX
 OS Bos taurus.
 XX
 PN US5973120-A.
 XX
 PD 26-OCT-1999.
 XX
 PF 07-MAR-1995; 95US-00399889.
 XX
 PR 30-NOV-1990; 90US-00621091.
 XX
 PA (UYVA) UNIV YALE.
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 PI Hudson BG, Reiders ST, Morrison KE;
 XX
 DR WPI; 1999-610317/52.
 DR N-PSDB; AA228774.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for
 PT diagnosis and treatment of Goodpasture syndrome.
 XX
 XX Claim 1; Col 31-34; 27pp; English.
 XX
 XX This sequence represents a recombinant bovine alpha3 chain of type IV
 CC collagen polypeptide. The sequence corresponds to the 238 amino acids of
 CC the C-terminal end of the triple helical domain and all 233 amino acids
 CC of the C-terminal non-collagenous domain. Alpha3 chain collagen
 CC polypeptides are useful for detecting Goodpasture antibodies in blood or
 CC tissue from a human patient and for treating Goodpasture syndrome,
 CC especially by neutralising the antibodies in the blood. The polypeptides
 CC also have a nephrotrophic activity
 XX
 XX Sequence 471 AA;

Query Match 90.3%; Score 1210.5; DB 2; Length 471;
 Best Local Similarity 90.6%; Pred. No. 3.6e-119;
 Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;
 QY 1 GLKGRGDSGSPATWTT-RGFVTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 59
 DB 227 GLKGRGDTGPPAAGAVMRGFFTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 286
 QY 60 DLGTLGSLQRFVTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 119
 DB 287 DLGTLGSLQRFVTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 346
 QY 120 SRTVCEGPAIAIVHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 179
 DB 347 SRTVCEGPAIAIVHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 406
 QY 180 EEFRAFPFIECHGRGTCTNYNSYSFWLASLNPMPKRPPISTVKAGELEKIIISRCQVC 239
 DB 407 EEFRAFPFIECHGRGTCTNYNSYSFWLASLNPMPKRPPISTVKAGELEKIIISRCQVC 466
 QY 240 MKKR 243
 DB 467 MKMR 470

RESULT 14
 AAY56783
 ID AAY56783 standard; protein; 471 AA.
 XX
 XX AAY56783;
 XX
 XX 27-MAR-2000 (first entry)
 XX
 XX Bovine alpha3 type IV collagen.
 XX
 XX Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.
 XX
 XX Bos sp.
 XX
 XX US6007980-A.
 XX
 XX 28-DEC-1999.
 XX
 XX 07-OCT-1998; 98US-00167364.
 XX
 XX 30-NOV-1990; 90US-00621091.
 XX
 XX 07-MAR-1995; 95US-00399889.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 XX (UYVA) UNIV YALE.
 XX
 XX Hudson BG, Reenders ST, Morrison KE;
 XX
 XX WPI; 2000-096371/08.
 XX
 XX N-PSDB; AAZ46728.

PT Diagnosing and treating Goodpasture syndrome using a peptide derived from
 PT type IV collagen.
 XX
 XX Disclosure; Col 19-24; 26pp; English.
 XX
 XX The invention provides a method of detecting Goodpasture antibodies in
 CC the fluid of a patient by contacting it with a peptide comprising at most
 CC 218 amino acids of the human alpha3 chain type IV collagen that contains
 CC the fragment shown in AAY56783. The methods are useful for the diagnosis
 CC and treatment of Goodpasture syndrome. The present sequence represents
 CC the bovine alpha3 chain of type IV collagen
 XX
 XX Sequence 471 AA;

Query Match 90.3%; Score 1210.5; DB 3; Length 471;
 Best Local Similarity 90.6%; Pred. No. 3.6e-119;
 Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;
 QY 1 GLKGRGDSGSPATWTT-RGFVTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 59
 DB 227 GLKGRGDTGPPAAGAVMRGFFTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 286
 QY 60 DLGTLGSLQRFVTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 119
 DB 287 DLGTLGSLQRFVTRHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 346
 QY 120 SRTVCEGPAIAIVHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 179
 DB 347 SRTVCEGPAIAIVHSQTTPAIPSCPGTVPVLYSGSFVQGNQAHGQ 406
 QY 180 EEFRAFPFIECHGRGTCTNYNSYSFWLASLNPMPKRPPISTVKAGELEKIIISRCQVC 239
 DB 407 EEFRAFPFIECHGRGTCTNYNSYSFWLASLNPMPKRPPISTVKAGELEKIIISRCQVC 466
 QY 240 MKKR 243
 DB 467 MKMR 470

RESULT 15
 AAE09483
 ID AAE09483 standard; protein; 471 AA.
 XX
 XX AAE09483;
 XX
 XX 19-NOV-2001 (first entry)
 XX
 XX Bovine alpha-3 chain of type IV collagen protein.
 XX
 XX Bovine; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
 XX Goodpasture syndrome.
 XX
 XX Bos taurus.
 XX
 XX US6277558-B1.
 XX
 XX 21-AUG-2001.
 XX
 XX 12-NOV-1999; 99US-00439897.
 XX
 XX 30-NOV-1990; 90US-00621091.
 XX
 XX 07-MAR-1995; 95US-00399889.
 XX
 XX 07-OCT-1998; 98US-00167364.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 XX Hudson BG;
 XX
 XX WPI; 2001-540401/60.
 XX
 XX N-PSDB; AAD16399.

Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
 PT Goodpasture antibodies from bodily fluid/tissue from patient or for

PT treating Goodpasture syndrome by contacting bodily fluid or tissue with
PT the polypeptide.

XX
PS Example 4; Col 33-36; 46pp; English.

XX
CC The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is bovine alpha-3 chain of type IV collagen protein

XX
SQ Sequence 471 AA;

Query Match 90.3%; Score 1210.5; DB 4; Length 471;
Best Local Similarity 90.6%; Pred. No. 3.6e-119;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;
QY 1 GLKGRGDSGPATWTT-RGFVETRHSTQTAIPSCPEGTVPLYSGFSLFVQGNQRAHQ 59
Db 227 GLKGRGDTGPPAAGAVMRGTFVTRHSQTAIPSCPEGTEPLYSGFSLFVQNEQAHGQ 286
QY 60 DLGTLGSLQRFITMPFLFCNVNDVCNFAVRNDYSYWLSTPALMPNMAPITGRALEPYI 119
Db 287 DLGTLGSLQRFITMPFLFCNVNDVCNFAVRNDYSYWLSTPALMPNMAPITGRALEPYI 346
QY 120 SRTVCEGPAIAVAHSQTTDIPCPHGWISLWKGSFIMFTSAGSEGTQALASPGSCL 179
Db 347 SRTVCEGPAIAVAHSQTTDIPCPHGWISLWKGSFIMFTSAGSEGTQALASPGSCL 406
QY 180 EEFRAFPFLECHGRGTCNYSNSYSFWLASLNPFRMRKPIPSTVKAGELEKIISRCQVC 239
Db 407 EEFRAFPFLECHGRGTCNYSNSYSFWLASLNPFRMRKPIPSTVKAGELEKIISRCQVC 466
QY 240 MKXR 243
Db 467 MKMR 470

Search completed: April 5, 2004, 06:58:30
Job time : 221.777 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 153.017 Seconds
(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-10
Perfect score: 1340
Sequence: 1 GLKRGKDGSSPATWTRGF.....KAGELEKIISRQVCVKKEH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 1340 | 100.0 | 244 | 14 | US-10-032-221B-10 |
| 2 | 1271 | 94.9 | 232 | 14 | US-10-032-221B-10 |
| 3 | 1052 | 78.5 | 191 | 14 | US-10-032-221B-22 |
| 4 | 960 | 71.6 | 406 | 9 | US-09-925-302-507 |
| 5 | 960 | 71.6 | 1669 | 15 | US-10-372-683-8 |
| 6 | 939 | 70.1 | 229 | 14 | US-10-032-221B-2 |
| 7 | 939 | 70.1 | 229 | 14 | US-10-032-221B-2 |
| 8 | 937.5 | 70.0 | 309 | 9 | US-09-925-297-496 |
| 9 | 928 | 69.3 | 229 | 14 | US-10-032-221B-306 |
| 10 | 880 | 65.7 | 211 | 14 | US-10-270-877-46 |
| 11 | 880 | 65.7 | 211 | 14 | US-10-270-877-46 |
| 12 | 852.5 | 63.6 | 1744 | 15 | US-10-369-493-5832 |
| 13 | 783.5 | 58.5 | 1759 | 15 | US-10-369-493-5832 |
| 14 | 755.5 | 56.4 | 430 | 9 | US-09-925-302-518 |
| 15 | 755.5 | 56.4 | 459 | 15 | US-10-331-496A-27 |

| | | | | | | |
|----|-------|------|------|----|---------------------|-------------------|
| 15 | 755.5 | 56.4 | 459 | 15 | US-10-372-683-30 | Sequence 30, Appl |
| 17 | 755.5 | 56.4 | 1712 | 10 | US-09-861-403-9 | Sequence 9, Appl |
| 18 | 741.5 | 55.3 | 227 | 14 | US-10-032-221B-6 | Sequence 303, App |
| 19 | 741.5 | 55.3 | 227 | 14 | US-10-032-221B-6 | Sequence 6, Appl |
| 20 | 727 | 54.3 | 228 | 14 | US-10-032-221B-23 | Sequence 307, App |
| 21 | 721 | 53.8 | 132 | 14 | US-10-032-221B-23 | Sequence 23, Appl |
| 22 | 690.5 | 53.5 | 231 | 14 | US-10-032-221B-23 | Sequence 305, App |
| 23 | 678 | 50.6 | 124 | 14 | US-10-032-221B-20 | Sequence 20, Appl |
| 24 | 662 | 49.4 | 120 | 14 | US-10-032-221B-21 | Sequence 21, Appl |
| 25 | 619 | 46.2 | 112 | 14 | US-10-032-221B-33 | Sequence 33, Appl |
| 26 | 480 | 35.8 | 88 | 14 | US-10-032-221B-33 | Sequence 34, Appl |
| 27 | 472.5 | 35.3 | 143 | 14 | US-10-032-221B-34 | Sequence 34, Appl |
| 28 | 471 | 35.1 | 88 | 14 | US-10-032-221B-25 | Sequence 25, Appl |
| 29 | 433 | 32.3 | 79 | 14 | US-10-032-221B-25 | Sequence 25, Appl |
| 30 | 353 | 26.3 | 64 | 14 | US-10-032-221B-26 | Sequence 26, Appl |
| 31 | 337 | 25.1 | 142 | 9 | US-09-864-761-38021 | Sequence 38021, A |
| 32 | 334 | 24.9 | 68 | 14 | US-10-270-877-50 | Sequence 50, Appl |
| 33 | 334 | 24.9 | 68 | 14 | US-10-270-877-50 | Sequence 50, Appl |
| 34 | 334 | 24.9 | 72 | 14 | US-10-270-877-48 | Sequence 48, Appl |
| 35 | 334 | 24.9 | 72 | 14 | US-10-270-877-48 | Sequence 52, Appl |
| 36 | 334 | 24.9 | 72 | 14 | US-10-270-877-52 | Sequence 48, Appl |
| 37 | 334 | 24.9 | 72 | 14 | US-10-270-877-52 | Sequence 52, Appl |
| 38 | 328 | 24.5 | 72 | 14 | US-10-270-877-61 | Sequence 61, Appl |
| 39 | 328 | 24.5 | 72 | 14 | US-10-270-877-61 | Sequence 61, Appl |
| 40 | 260 | 19.4 | 70 | 11 | US-09-864-408A-1258 | Sequence 1258, Ap |
| 41 | 201 | 15.0 | 46 | 9 | US-09-864-761-48095 | Sequence 48095, A |
| 42 | 189.5 | 14.1 | 70 | 9 | US-09-864-761-37448 | Sequence 37448, A |
| 43 | 189.5 | 14.1 | 70 | 9 | US-09-864-761-47938 | Sequence 47938, A |
| 44 | 147 | 11.0 | 25 | 14 | US-10-032-221B-37 | Sequence 37, Appl |
| 45 | 146 | 10.9 | 27 | 14 | US-10-032-221B-39 | Sequence 39, Appl |

ALIGNMENTS

RESULT 1

US-10-032-221B-10
; Sequence 10, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREO
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-10

Query Match 100.0%; Score 1340; DB 14; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.1e-132;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKRGKDGSSPATWTRGFVTRHSQTATPSCSEGVPLYSRFLVQGNQRAHQD 60

Db 1 GLKGRGDSGPATWTTTRGVFTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQD 60
QY 61 LGTLGSLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGRALPEYIS 120
Db 61 LGTLGSLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGRALPEYIS 120
QY 121 RCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
Db 121 RCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180
QY 181 EFRASPLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCM 240
Db 181 EFRASPLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCM 240
QY 241 KKRH 244
Db 241 KKRH 244

RESULT 2

US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 304
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 3 chain
US-10-206-699-304

Query Match 94.9%; Score 1271; DB 14; Length 232;
Best Local Similarity 99.6%; Pred. No. 2e-125;
Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 13 ATWTTTRGVFTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLSCLQRF 72
Db 1 ATWTTTRGVFTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLSCLQRF 60
QY 73 TMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGRALPEYISRCTVCEGPAIAI 132
Db 61 TMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGRALPEYISRCTVCEGPAIAI 120
QY 133 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLEEFASPFLECHG 192
Db 121 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLEEFASPFLECHG 180
QY 193 RGTCTNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCMKKRH 244
Db 181 RGTCTNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCMKKRH 232

RESULT 3

US-10-032-221B-22
; Sequence 22, Application US/10032221B

Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match 78.5%; Score 1052; DB 14; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.8e-102;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 54 QRAHGQDGLTGLSCLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGR 113
Db 1 QRAHGQDGLTGLSCLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGR 60
QY 114 ALEPYISRCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALA 173
Db 61 ALEPYISRCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALA 120
QY 174 SPGSCLEEFASPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKII 233
Db 121 SPGSCLEEFASPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKII 180
QY 234 SRCQVCMKKRH 244
Db 181 SRCQVCMKKRH 191

RESULT 4

US-09-925-302-507
; Sequence 507, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: PCT/US00/05918
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 507
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

NAME/KEY: SITE
LOCATION: (71)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-507

Query Match 71.6%; Score 960; DB 9; Length 406;
Best Local Similarity 69.0%; Pred. No. 2.4e-92;
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKGRGDSGPATWTRGVFTRHSGQTTPAIPSCPGTVPVLYSGFSLVQGNQRAHQD 60
DB 166 GLPGSMGPPGTPS--VDHGFVLRHSGTIDDPQCPGSGTKILYHGYSLLYVQGNRAHQD 223
QY 61 LGTLGSLQSFTHMPFLFCNVNDVCFASNDYSYWLSTPALMPMNAIPITGRALPEYIS 120
DB 224 LGTAGSLRKFSTMPFLFCNVNDVCFASNDYSYWLSTPEPMPMNAIPITGRALPEYIS 283
QY 121 RCTVCEGPAIAVHSGTTPDPCPHGWISLWKGFSFIMFTSAGSGTQOALASPGSCLE 180
DB 284 RCACEAPAMVAVHSGTTPDPCPGSWSLWIGYSFVHVSAGSGOALASPGSCLE 343
QY 181 EFRASPLECHGRGTCNYNSYSFVWLASLNPMPKPIPTSVKAGELEKIISRCQVCM 240
DB 344 EFRASPLECHGRGTCNYNSYSFVWLASLNPMPKPIPTSVKAGELEKIISRCQVCM 403
QY 241 KK 242
DB 404 RR 405

RESULT 5
US-10-372-683-8
Sequence 8, Application US/10372683
Publication No. US20040009171A1
GENERAL INFORMATION:
APPLICANT: GERRITSEN, MARY E.
APPLICANT: PEARLE JR., FRANKLIN V.
TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
FILE REFERENCE: P1928R1P1
CURRENT APPLICATION NUMBER: US/10/372,683
CURRENT FILING DATE: 2003-02-21
PRIOR APPLICATION NUMBER: US 10/271,690
PRIOR FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: US 60/344,534
PRIOR FILING DATE: 2001-10-18
NUMBER OF SEQ ID NOS: 49
SEQ ID NO 8
LENGTH: 1669
TYPE: PRT
ORGANISM: Homo sapien
US-10-372-683-8

Query Match 71.6%; Score 960; DB 15; Length 1669;
Best Local Similarity 69.0%; Pred. No. 1.5e-91;
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKGRGDSGPATWTRGVFTRHSGQTTPAIPSCPGTVPVLYSGFSLVQGNQRAHQD 60
DB 1429 GLPGSMGPPGTPS--VDHGFVLRHSGTIDDPQCPGSGTKILYHGYSLLYVQGNRAHQD 1486
QY 61 LGTLGSLQSFTHMPFLFCNVNDVCFASNDYSYWLSTPALMPMNAIPITGRALPEYIS 120
DB 1487 LGTAGSLRKFSTMPFLFCNVNDVCFASNDYSYWLSTPEPMPMNAIPITGRALPEYIS 1546
QY 121 RCTVCEGPAIAVHSGTTPDPCPHGWISLWKGFSFIMFTSAGSGTQOALASPGSCLE 180
DB 1547 RCACEAPAMVAVHSGTTPDPCPGSWSLWIGYSFVHVSAGSGOALASPGSCLE 1606
QY 181 EFRASPLECHGRGTCNYNSYSFVWLASLNPMPKPIPTSVKAGELEKIISRCQVCM 240
DB 1607 EFRASPLECHGRGTCNYNSYSFVWLASLNPMPKPIPTSVKAGELEKIISRCQVCM 1666

QY 241 KK 242
DB 1667 RR 1668

RESULT 6
US-10-206-699-302
Sequence 302, Application US/10206699
Publication No. US20030100510A1
GENERAL INFORMATION:
APPLICANT: Sundaramoorthy, M.
APPLICANT: Hudson, B.
TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
FILE REFERENCE: MBHB 01-1017
CURRENT APPLICATION NUMBER: US/10/206,699
CURRENT FILING DATE: 2002-07-26
PRIOR APPLICATION NUMBER: US 60/308,523
PRIOR FILING DATE: 2001-07-27
PRIOR APPLICATION NUMBER: US 60/351,289
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: US 60/366,854
PRIOR FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: US 60/385,362
PRIOR FILING DATE: 2002-06-03
NUMBER OF SEQ ID NOS: 307
SOFTWARE: Patent in version 3.1
SEQ ID NO 302
LENGTH: 229
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: alpha 1 chain
US-10-206-699-302

Query Match 70.1%; Score 939; DB 14; Length 229;
Best Local Similarity 71.9%; Pred. No. 1.9e-90;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTPAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLQRFTTMEPLF 78
DB 5 GFLVTRHSQTIDDPQCPGSGTKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTMEPLF 64
QY 79 CNVNDVCFASNDYSYWLSTPALMPMNAIPITGRALPEYISRCTVCEGPAIAVHSGT 138
DB 65 CNINNVCNFAASNDYSYWLSTPEPMPMNAIPITGRALPEYISRCTVCEGPAIAVHSGT 124
QY 139 TDIPPCPHGWISLWKGFSFIMFTSAGSGTQOALASPGSCLEFRASPLECHGRGTCNY 198
DB 125 IQIPPCPGSWSLWIGYSFVHVSAGSGOALASPGSCLEFRASPLECHGRGTCNY 184
QY 199 YSNYSYFVWLASLNPMPKPIPTSVKAGELEKIISRCQVCMKK 242
DB 185 YANAYSPFLATIERSEMFKKPTPTLTKAGELRTHVSRQVCMRR 228

RESULT 7
US-10-032-221B-2
Sequence 2, Application US/10032221B
Publication No. US20030144481A1
GENERAL INFORMATION:
APPLICANT: Kalluri, Raghuram
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
FILE REFERENCE: 2312/20082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118

;; PRIOR FILING DATE: 2000-01-07
;; PRIOR APPLICATION NUMBER: US 09/335,224
;; PRIOR FILING DATE: 1999-06-17
;; PRIOR APPLICATION NUMBER: US 60/126,175
;; PRIOR FILING DATE: 1999-03-25
;; PRIOR APPLICATION NUMBER: US 60/089,689
;; PRIOR FILING DATE: 1998-06-17
;; NUMBER OF SEQ ID NOS: 58
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 2
;; LENGTH: 229
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-032-221B-2

Query Match 70.1%; Score 939; DB 14; Length 229;
Best Local Similarity 71.9%; Pred. No. 1.9e-90;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;
QY 19 GFVETRHSTQTAIPSCPEGTVPLVSGFSLFVQGNORAHGQDLGTGSCLOREFTTTPFLF 78
DB 5 GFVETRHSTQTDPPQPSGKILYHGYSLLVQGNERAGQDLGTAGSCLRFKSTMPFLF 64
QY 79 CNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQT 138
DB 65 CNINNVCFASRNDYSYWLSTPEPMPMSAPITGENIRPFISRCVACEAPAMVMAVHSQT 124
QY 139 TDTPCPHGWSLWKGFSTMTFTSAGSEGTGQALASPGSCLEEFASPELECHGRGTGNY 198
DB 125 IQTPPCFSGWSLWIGYFVWHTSAGAGSGQALASPGSCLEEFASPELECHGRGTGNY 184
QY 199 YNSYSFSLASLNPFRMRKPISTVKAGELEKIIISRCQVCMKK 242
DB 185 YANYSFWLATIERSEWFKKPTSTLKAGELRTHVSRQCQVCMKR 228

RESULT 8
US-09-925-297-496
;; Sequence 496, Application US/09925297
;; Patent No. US20020081659A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;; FILE REFERENCE: PA105
;; CURRENT APPLICATION NUMBER: US/09/925,297
;; CURRENT FILING DATE: 2001-08-10
;; PRIOR APPLICATION NUMBER: PCT/US00/05989
;; PRIOR FILING DATE: 2000-03-08
;; PRIOR APPLICATION NUMBER: 60/124,270
;; PRIOR FILING DATE: 1999-03-12
;; NUMBER OF SEQ ID NOS: 928
;; SOFTWARE: Patent in Ver. 2.0
;; SEQ ID NO 496
;; LENGTH: 309
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (247)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-496

Query Match 70.0%; Score 937.5; DB 9; Length 309;
Best Local Similarity 67.5%; Pred. No. 3.9e-90;
Matches 164; Conservative 35; Mismatches 43; Indels 1; Gaps 1;
QY 1 GLKGRDSDSGSPATWT-TRGFVETRHSTQTAIPSCPEGTVPLVSGFSLFVQGNORAHGQ 59
DB 66 GPDGLQGPFPPTGSSVAHGFILTRHSQTTDAPCCQGTQVYEGFSLVQGNKRAHQ 125
QY 60 DLGTGSCLOREFTTTPFLFONVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIYI 119
DB 126 DLGTAGSCLRRFSTMPFMCNINNVCFASRNDYSYWLSTPEPMPMSAPITGENIRPFIS 185

QY 120 SRTVCEGPAIAIAVHSQTTDTPCPHGWSLWKGFSTMTFTSAGSEGTGQALASPGSCL 179
DB 186 SRCVACEAPAVIAVHSQTIQPHCQGWDSLWIGYSFMMHTSAGAGSGQALASPGSCL 245
QY 180 EEFASPFLECHGRGTGTCNYSNSYSFWLASLNPFRMRKPISTVKAGELEKIIISRCQVC 239
DB 246 EEFASPFLECHGRGTGTCNYSNSYSFWLASLNPFRMRKPISTVKAGELEKIIISRCQVC 305
QY 240 MKK 242
DB 306 MKR 308
RESULT 9
US-10-206-699-306
;; Sequence 306, Application US/10206699
;; Publication No. US20030100510A1
;; GENERAL INFORMATION:
;; APPLICANT: Sundaramoorthy, M.
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
;; FILE REFERENCE: MHHB 01-1017
;; CURRENT APPLICATION NUMBER: US/10/206,699
;; CURRENT FILING DATE: 2002-07-26
;; PRIOR APPLICATION NUMBER: US 60/308,523
;; PRIOR FILING DATE: 2001-07-27
;; PRIOR APPLICATION NUMBER: US 60/351,289
;; PRIOR FILING DATE: 2001-10-29
;; PRIOR APPLICATION NUMBER: US 60/366,854
;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 306
;; LENGTH: 229
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; OTHER INFORMATION: alpha 5 chain
US-10-206-699-306

Query Match 69.3%; Score 928; DB 14; Length 229;
Best Local Similarity 71.0%; Pred. No. 2.7e-89;
Matches 159; Conservative 33; Mismatches 32; Indels 0; Gaps 0;
QY 19 GFVETRHSTQTAIPSCPEGTVPLVSGFSLFVQGNORAHGQDLGTGSCLOREFTTTPFLF 78
DB 5 GFVETRHSTQTDAPCCQGTQVYEGFSLVQGNKRAHQDLGTAGSCLRRSTWPFMF 64
QY 79 CNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQT 138
DB 65 CNINNVCFASRNDYSYWLSTPEPMPMSAPITGENIRPFISRCVACEAPAVIAVHSQT 124
QY 139 TDTPCPHGWSLWKGFSTMTFTSAGSEGTGQALASPGSCLEEFASPELECHGRGTGNY 198
DB 125 IQTPPCFSGWSLWIGYFVWHTSAGAGSGQALASPGSCLEEFASPELECHGRGTGNY 184
QY 199 YNSYSFSLASLNPFRMRKPISTVKAGELEKIIISRCQVCMKK 242
DB 185 YANYSFWLATVDVSDMFSKPSQSETLKAGDLRTRISRCQVCMKR 228

RESULT 10
US-10-270-877-46
;; Sequence 46, Application US/10270877
;; Publication No. US20030049791A1
;; GENERAL INFORMATION:
;; APPLICANT: Saus, Juan
;; TITLE OF INVENTION: Goodpasture Binding Protein
;; FILE REFERENCE: 98-723-AD1


```

; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match          65.7%; Score 880; DB 14; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.7e-84;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Db 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSCLOQFTTTPFLFCNVNDVCFASRNDSYWLSTPALMPNMNAPITGRALEPYIS 120
Db 61 LGTLGSCLOQFTTTPFLFCNVNDVCFASRNDSYWLSTPALMPNMNAPITGRALEPYIS 120
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFIM 159
Db 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFIM 159

RESULT 11
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match          65.7%; Score 880; DB 14; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.7e-84;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Db 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSCLOQFTTTPFLFCNVNDVCFASRNDSYWLSTPALMPNMNAPITGRALEPYIS 120
Db 61 LGTLGSCLOQFTTTPFLFCNVNDVCFASRNDSYWLSTPALMPNMNAPITGRALEPYIS 120
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFIM 159
Db 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFIM 159

```

```

RESULT 12
US-10-369-493-5832
; Sequence 5832, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5832
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5832

Query Match          63.6%; Score 852.5; DB 15; Length 1744;
Best Local Similarity 60.7%; Pred. No. 3.4e-80;
Matches 148; Conservative 35; Mismatches 60; Indels 1; Gaps 1;

QY 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQ 59
Db 1501 GLPGTGPYSPGOWAPSRGFTFAKHSQTTAVPQCPPGASQLWEGYSLLYVQNGRASGQ 1560
QY 60 DLGLGSCLOQFTTTPFLFCNVNDVCFASRNDSYWLSTPALMPNMNAPITGRALEPYI 119
Db 1561 DLGQPGSCLSKFNTPMFCNMNSVCHVSRNDSYWLSTDEPTMMPMPVTTGTAIRPYI 1620
QY 120 SRCVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFIMFTSAGSEGTQALASPGSCL 179
Db 1621 SRCACVBPQTIIAVHSQDTSVPQCPQGSQMTGYSFVMTAAGAGTQSLQSPGSC 1680
QY 180 EEFRAFPFLECHRGCTGCTNYNSYSYFWLASLNPERMFKPIPTVKAGELEKIIISRCOV 239
Db 1681 EEFRAVPFIECHRGCTGCTNYATNHGFWSLIVDQDKFRKPMSTLXAGGLKDRVSRQVC 1740
QY 240 MKXR 243
Db 1741 LKXR 1744

RESULT 13
US-10-369-493-7032
; Sequence 7032, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7032
; LENGTH: 1759
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-7032

```

Query Match 58.5%; Score 783.5; DB 15; Length 1759;
Best Local Similarity 55.3%; Pred. No. 6.4e-73;
Matches 142; Conservative 41; Mismatches 59; Indels 15; Gaps 4;

QY 1 GLKGRGDSGSP-----ATWTRGVFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQ 51
DB 1505 GLDQGGGFGAGLPGAPGAAGPAVDGFLVKHSQTTVEVPCPGQTKLWDGYSLLXIE 1564

QY 52 GQRAHQDGLTGLSCLOQFTTMPELFCNNVNDVCFNPNRNDYSYWLSTPALMPMNPAPIT 111
DB 1565 GNEKSHNQLGHAGSCLOQFTTMPELFCNNVNDVCFNPNRNDYSYWLSTSEALP--MMPVN 1622

QY 112 GRALEYISRCTVCEGPAIAVHSQTTDIPPCPHGWSLWKGFSFIMFTSAGSGTGQA 171
DB 1623 EREIEYISRCVCAVECAPANTIAVHSQTTQIPNCPAGWSLWIGYFAMHTGAGGGQOS 1682

QY 172 LASPGSCLEEFRAFPFLECHG-RGTCNNYSNSYFWLASLNPERMFRKPIPTVKAGBLE 230
DB 1683 LSPSGSCLEDFRATPFIECNARGSGCHYFANKFSFWLTTIDNDSFKVPESQTLKGNLR 1742

QY 231 KIISRCOVCMK---RH 244
DB 1743 TRVSRQVCVKSTPDRH 1759

RESULT 14
US-09-925-302-518
; Sequence 518, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 518
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-518

Query Match 56.4%; Score 755.5; DB 9; Length 430;
Best Local Similarity 58.0%; Pred. No. 9.1e-71;
Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;

QY 1 GLKGRGDSGSPATWTRGVFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60
DB 189 GRPGSPGLPGMGSRVSGIGLLVKHSQTDQEPKCPVGMNKLWSGYSLLYFEGQEKAHNQD 248

QY 61 LGTGLSCLOQFTTMPELFCNNVNDVCFNPNRNDYSYWLSTPALMPMNPAPITGRALPEYIS 120
DB 249 LGLAGSCLARFSTMPFLYCNPGDVCIYASRNDKSYWLSTTA--PLPMPVAEDEIKPYIS 306

QY 121 RCTVCEGPAIAVHSQTTDIPPCPHGWSLWKGFSFIMFTSAGSGTGQAALASPGSCLE 180
DB 307 RCSVCEAPAIATAVHSQDVSIPHCPCAGWSLWIGYSFLMHTAAGDEGGQSLVSPGSCLE 366

QY 181 EFRASPFLECH-GRGTCNNYSNSYFWLASLNPERMFR-KPIPTVKAGLEKIIISRCOV 238
DB 367 DFRATPFIECNNGRGTCCHYANKYSFWLTTI-PEQSQGSPSADTLKAGLIRTHISRCOV 425

QY 239 CMK 241

DB 426 CMK 428

RESULT 15
US-10-331-496A-27
; Sequence 27, Application US/10331496A
; Publication No. US20030228305A1
; GENERAL INFORMATION:
; APPLICANT: FRANTZ, GRETCHEN
; APPLICANT: HILLAN, KENNETH J.
; APPLICANT: PHILLIPS, HEIDI S.
; APPLICANT: POLAKIS, PAUL
; APPLICANT: SMITH, VICTORIA
; APPLICANT: SPENCER, SUSAN D.
; APPLICANT: WILLIAMS, P. MICKEY
; APPLICANT: WU, THOMAS D.
; APPLICANT: ZHANG, ZEMIN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF TUMOR
; FILE REFERENCE: P5014R1-PCT
; CURRENT APPLICATION NUMBER: US/10/331,496A
; CURRENT FILING DATE: 2002-12-30
; PRIOR FILING DATE: 2002-01-02 US 60/345,444
; PRIOR APPLICATION NUMBER: US 60/351,885
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/360,066
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: US 60/362,004
; PRIOR FILING DATE: 2002-03-05 US 60/366,869
; PRIOR FILING DATE: 2002-03-20 US 60/366,284
; PRIOR FILING DATE: 2002-03-21 US 60/368,679
; PRIOR FILING DATE: 2002-03-28 US 60/404,809
; PRIOR FILING DATE: 2002-08-19 US 60/405,645
; PRIOR FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 27
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-331-496A-27

Query Match 56.4%; Score 755.5; DB 15; Length 459;
Best Local Similarity 58.0%; Pred. No. 1e-70;
Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;

QY 1 GLKGRGDSGSPATWTRGVFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60
DB 218 GRPGSPGLPGMGSRVSGIGLLVKHSQTDQEPKCPVGMNKLWSGYSLLYFEGQEKAHNQD 277

QY 61 LGTGLSCLOQFTTMPELFCNNVNDVCFNPNRNDYSYWLSTPALMPMNPAPITGRALPEYIS 120
DB 278 LGLAGSCLARFSTMPFLYCNPGDVCIYASRNDKSYWLSTTA--PLPMPVAEDEIKPYIS 335

QY 121 RCTVCEGPAIAVHSQTTDIPPCPHGWSLWKGFSFIMFTSAGSGTGQAALASPGSCLE 180
DB 336 RCSVCEAPAIATAVHSQDVSIPHCPCAGWSLWIGYSFLMHTAAGDEGGQSLVSPGSCLE 395

QY 181 EFRASPFLECH-GRGTCNNYSNSYFWLASLNPERMFR-KPIPTVKAGLEKIIISRCOV 238
DB 396 DFRATPFIECNNGRGTCCHYANKYSFWLTTI-PEQSQGSPSADTLKAGLIRTHISRCOV 454

QY 239 CMK 241
DB 455 CMK 457

Mon Apr 5 07:53:04 2004

us-10-032-221b-10.rapb

Page 7

Search completed: April 5, 2004, 07:36:05
Job time : 154.017 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 57.3075 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221b-10
Perfect score: 1340
Sequence: 1 GLKRGDGSPTWTRGP.....KAGELEKLSRCQVCMKKRH 244

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 1314 | 98.1 | 268 | 4 | US-09-589-927-6 |
| 2 | 1314 | 98.1 | 268 | 4 | US-09-277-665-6 |
| 3 | 1314 | 98.1 | 268 | 4 | US-09-589-987-6 |
| 4 | 1210.5 | 90.3 | 471 | 2 | US-08-399-889-24 |
| 5 | 1210.5 | 90.3 | 471 | 3 | US-09-167-364-24 |
| 6 | 1210.5 | 90.3 | 471 | 3 | US-09-439-897-2 |
| 7 | 1192 | 89.0 | 218 | 2 | US-08-399-889-25 |
| 8 | 1192 | 89.0 | 218 | 3 | US-09-167-364-25 |
| 9 | 1192 | 89.0 | 218 | 3 | US-09-439-897-4 |
| 10 | 939 | 70.1 | 260 | 4 | US-09-589-927-2 |
| 11 | 939 | 70.1 | 260 | 4 | US-09-277-665-2 |
| 12 | 939 | 70.1 | 260 | 4 | US-09-589-987-2 |
| 13 | 938.5 | 70.0 | 264 | 4 | US-09-589-927-10 |
| 14 | 938.5 | 70.0 | 264 | 4 | US-09-277-665-10 |
| 15 | 938.5 | 70.0 | 264 | 4 | US-09-589-987-10 |
| 16 | 880 | 65.7 | 211 | 4 | US-09-512-563C-46 |
| 17 | 741.5 | 55.3 | 258 | 4 | US-09-589-927-4 |
| 18 | 741.5 | 55.3 | 258 | 4 | US-09-277-665-4 |
| 19 | 741.5 | 55.3 | 258 | 4 | US-09-589-987-4 |
| 20 | 727 | 54.3 | 260 | 4 | US-09-589-927-12 |
| 21 | 727 | 54.3 | 260 | 4 | US-09-277-665-12 |
| 22 | 727 | 54.3 | 260 | 4 | US-09-589-987-12 |
| 23 | 690.5 | 51.5 | 260 | 4 | US-09-589-927-8 |
| 24 | 690.5 | 51.5 | 260 | 4 | US-09-277-665-8 |
| 25 | 690.5 | 51.5 | 260 | 4 | US-09-589-987-8 |
| 26 | 339.5 | 25.3 | 1694 | 1 | US-08-494-168-2 |
| 27 | 334 | 24.9 | 68 | 4 | US-09-512-563C-50 |

| | | | | | | |
|----|------|------|-----|---|--------------------|--------------------|
| 28 | 334 | 24.9 | 72 | 4 | US-09-512-563C-48 | Sequence 48, Appl |
| 29 | 334 | 24.9 | 72 | 4 | US-09-512-563C-52 | Sequence 52, Appl |
| 30 | 328 | 24.5 | 72 | 4 | US-09-512-563C-61 | Sequence 61, Appl |
| 31 | 189 | 14.1 | 36 | 3 | US-09-439-897-65 | Sequence 65, Appl |
| 32 | 143 | 10.7 | 26 | 3 | US-09-439-897-63 | Sequence 63, Appl |
| 33 | 116 | 8.7 | 21 | 4 | US-09-512-563C-26 | Sequence 26, Appl |
| 34 | 113 | 8.4 | 21 | 4 | US-09-512-563C-27 | Sequence 27, Appl |
| 35 | 110 | 8.2 | 35 | 3 | US-09-439-897-64 | Sequence 64, Appl |
| 36 | 98 | 7.3 | 15 | 3 | US-09-439-897-61 | Sequence 61, Appl |
| 37 | 85.5 | 6.4 | 409 | 4 | US-09-198-452A-554 | Sequence 554, App |
| 38 | 83.5 | 6.2 | 404 | 1 | US-07-602-824A-4 | Sequence 4, Appl |
| 39 | 83.5 | 6.2 | 404 | 1 | US-07-602-608-4 | Sequence 4, Appl |
| 40 | 83.5 | 6.2 | 404 | 1 | US-07-983-451-4 | Sequence 4, Appl |
| 41 | 83.5 | 6.2 | 404 | 1 | US-08-261-578-4 | Sequence 4, Appl |
| 42 | 83.5 | 6.2 | 404 | 1 | US-08-261-577-10 | Sequence 10, Appl |
| 43 | 83.5 | 6.2 | 539 | 6 | 5449756-4 | Patent No. 5449756 |
| 44 | 82 | 6.1 | 15 | 3 | US-09-439-897-53 | Sequence 53, Appl |
| 45 | 82 | 6.1 | 15 | 3 | US-09-439-897-57 | Sequence 57, Appl |

ALIGNMENTS

RESULT 1
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match 98.1%; Score 1314; DB 4; Length 268;
Best Local Similarity 99.6%; Pred. No. 8.4e-138;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 5 | KRGDSGSPATWTTTRGFVTRHSQTATPSCPEGTVPLYSGFSLFVQGNORAHGQDLGTL | 64 |
| DB | 29 | KRGDSGSPATWTTTRGFVTRHSQTATPSCPEGTVPLYSGFSLFVQGNORAHGQDLGTL | 88 |
| QY | 65 | GSCLQRFTTTPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNAPIITGRALEPIYSRCTV | 124 |
| DB | 89 | GSCLQRFTTTPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNAPIITGRALEPIYSRCTV | 148 |
| QY | 125 | CEGPAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALAPGSCLEEFRA | 184 |
| DB | 149 | CEGPAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALAPGSCLEEFRA | 208 |
| QY | 185 | SPFLECHGRGTCNYNSYSYFWLASLNPENFRKPIPTVKAGLEKIIISRCQVCMKKRH | 244 |
| DB | 209 | SPFLECHGRGTCNYNSYSYFWLASLNPENFRKPIPTVKAGLEKIIISRCQVCMKKRH | 268 |

RESULT 2
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-I
; CURRENT APPLICATION NUMBER: US/09/277,665

```
/ CURRENT FILING DATE: 1999-03-26
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 268
/ TYPE: PRT
/ ORGANISM: Human
US-09-277-663-6

Query Match      98.1%; Score 1314; DB 4; Length 268;
Best Local Similarity 99.6%; Pred. No. 8.4e-138;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGDSGSPATWTRGFTVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQDLGTL 64
DB 29 KRGDSGSPATWTRGFTVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQDLGTL 88

QY 65 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMNMAPITGRALEPYISRCTV 124
DB 89 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMNMAPITGRALEPYISRCTV 148

QY 125 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFSLFIMFTSAGSEGTQALASPGSCLEEFRA 184
DB 149 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFSLFIMFTSAGSEGTQALASPGSCLEEFRA 208

QY 185 SPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCKKXKH 244
DB 209 SPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCKKXKH 268

RESULT 3
US-09-589-987-6
/ Sequence 6, Application US/0959987
/ Patent No. 6498140
/ GENERAL INFORMATION:
/ APPLICANT: University of Kansas Medical Center
/ TITLE OF INVENTION: The use of isolated domains of type IV collagen to
/ FILE REFERENCE: 945251
/ CURRENT APPLICATION NUMBER: US/09/589,987
/ CURRENT FILING DATE: 2000-06-07
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 268
/ TYPE: PRT
/ ORGANISM: Human
US-09-589-987-6

Query Match      98.1%; Score 1314; DB 4; Length 268;
Best Local Similarity 99.6%; Pred. No. 8.4e-138;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGDSGSPATWTRGFTVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQDLGTL 64
DB 29 KRGDSGSPATWTRGFTVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQDLGTL 88

QY 65 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMNMAPITGRALEPYISRCTV 124
DB 89 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMNMAPITGRALEPYISRCTV 148

QY 125 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFSLFIMFTSAGSEGTQALASPGSCLEEFRA 184
DB 149 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFSLFIMFTSAGSEGTQALASPGSCLEEFRA 208

QY 185 SPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCKKXKH 244
DB 209 SPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVCKKXKH 268

RESULT 4
US-08-399-889-24
/ Sequence 24, Application US/08399889B

/ Patent No. 5973120
/ GENERAL INFORMATION:
/ APPLICANT: Reeder, Stephen T
/ APPLICANT: Morrison, Karen E
/ APPLICANT: Hudson, Billy G
/ TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
/ FILE REFERENCE: 951263A
/ CURRENT APPLICATION NUMBER: US/08/399,889B
/ CURRENT FILING DATE: 1995-03-07
/ EARLIER APPLICATION NUMBER: 07/621091
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 24
/ LENGTH: 471
/ TYPE: PRT
/ ORGANISM: Calf
US-08-399-889-24

Query Match      90.3%; Score 1210.5; DB 3; Length 471;
Best Local Similarity 90.6%; Pred. No. 5.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRGDSGSPATWTT-RGFVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQ 59
DB 227 GLKGRGDSGSPATWTT-RGFVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQ 286

QY 60 DLGTLGSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMNMAPITGRALEPYI 119
DB 287 DLGTLGSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMNMAPITGRALEPYI 346

QY 120 SRTVCCEGPAIAIAVHSQTTDIPCPCHGWISLWKGFSLFIMFTSAGSEGTQALASPGSCL 179
DB 347 SRTVCCEGPAIAIAVHSQTTDIPCPCHGWISLWKGFSLFIMFTSAGSEGTQALASPGSCL 406

QY 180 EEFRASPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVC 239
DB 407 EEFRASPFLECHGRGTCNYNSYSFSLASLNPERMFRKPIPTSVKAGELEKIIISRCQVC 466

QY 240 MKKR 243
DB 467 MKKR 470

RESULT 5
US-09-167-364-24
/ Sequence 24, Application US/09167364
/ Patent No. 6007980
/ GENERAL INFORMATION:
/ APPLICANT: Reeder, Stephen T
/ APPLICANT: Morrison, Karen E
/ APPLICANT: Hudson, Billy G
/ TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
/ FILE REFERENCE: 951263B
/ CURRENT APPLICATION NUMBER: US/09/167,364
/ CURRENT FILING DATE: 1998-10-07
/ EARLIER APPLICATION NUMBER: 08/399889
/ CURRENT FILING DATE: 1995-03-07
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 24
/ LENGTH: 471
/ TYPE: PRT
/ ORGANISM: Calf
US-09-167-364-24

Query Match      90.3%; Score 1210.5; DB 3; Length 471;
Best Local Similarity 90.6%; Pred. No. 5.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRGDSGSPATWTT-RGFVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQ 59
DB 227 GLKGRGDSGSPATWTT-RGFVTRHSQTTPSCPEGTPLYSGFSLFVQGNQRAHGQ 286
```

QY 60 DLGTLGSLQRTTTPPFLFCNVNDVNCNFRNDYSYWLSTPALMPNMNMAPITGRALPEYI 119
DB 287 DLGTLGSLQRTTTPPFLFCNVNDVNCNFRNDYSYWLSTPALMPNMNMAPITGRALPEYI 346
QY 120 SRTCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGOALASPGSCL 179
DB 347 SRTCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGOALASPGSCL 406
QY 180 BEFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVC 239
DB 407 BEFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVC 466
QY 240 MKKR 243
DB 467 MKKR 470

RESULT 6

US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 90.3%; Score 1210.5; DB 3; Length 471;
Best Local Similarity 90.6%; Pred. No. 5.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;
QY 1 GLKKGDSRATWTT-RGFVTRHSQTTAIPSCPEGTVPYSGFSFLFVQGNQRAHQ 59
DB 227 GLKKGDSRATWTT-RGFVTRHSQTTAIPSCPEGTVPYSGFSFLFVQGNQRAHQ 286
QY 60 DLGTLGSLQRTTTPPFLFCNVNDVNCNFRNDYSYWLSTPALMPNMNMAPITGRALPEYI 119
DB 287 DLGTLGSLQRTTTPPFLFCNVNDVNCNFRNDYSYWLSTPALMPNMNMAPITGRALPEYI 346
QY 120 SRTCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGOALASPGSCL 179
DB 347 SRTCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGOALASPGSCL 406
QY 180 BEFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVC 239
DB 407 BEFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVC 466
QY 240 MKKR 243
DB 467 MKKR 470

RESULT 7

US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeder, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07

; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

Query Match 89.0%; Score 1192; DB 2; Length 218;
Best Local Similarity 99.5%; Pred. No. 2.3e-124;
Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 27 QTTAIPSCPEGTVPYSGFSFLFVQGNQRAHQDGLTGLSCLQRTTTPPFLFCNVNDVNCN 86
DB 1 QTTAIPSCPEGTVPYSGFSFLFVQGNQRAHQDGLTGLSCLQRTTTPPFLFCNVNDVNCN 60
QY 87 FASRNDYSYWLSTPALMPNMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALMPNMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120
QY 147 GWISLWKGFSPIMFTSAGSEGTGOALASPGSCLBEFRASPFLECHGRGTCNYNSYSYFW 206
DB 121 GWISLWKGFSPIMFTSAGSEGTGOALASPGSCLBEFRASPFLECHGRGTCNYNSYSYFW 180
QY 207 LASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCCKKRH 244
DB 181 LASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCCKKRH 218

RESULT 8

US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeder, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match 89.0%; Score 1192; DB 3; Length 218;
Best Local Similarity 99.5%; Pred. No. 2.3e-124;
Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 27 QTTAIPSCPEGTVPYSGFSFLFVQGNQRAHQDGLTGLSCLQRTTTPPFLFCNVNDVNCN 86
DB 1 QTTAIPSCPEGTVPYSGFSFLFVQGNQRAHQDGLTGLSCLQRTTTPPFLFCNVNDVNCN 60
QY 87 FASRNDYSYWLSTPALMPNMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALMPNMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120
QY 147 GWISLWKGFSPIMFTSAGSEGTGOALASPGSCLBEFRASPFLECHGRGTCNYNSYSYFW 206
DB 121 GWISLWKGFSPIMFTSAGSEGTGOALASPGSCLBEFRASPFLECHGRGTCNYNSYSYFW 180
QY 207 LASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCCKKRH 244
DB 181 LASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCCKKRH 218

```

156 IQIPPCSGWSSLWIGYSFVWHTSAGAGSGQALASPGSCLEBFRSAPFIECHGRGTCNY 215
199 YSNSYSFWLASLNPERMFRKPIPTSTVKAGELKLIISRCQVCMKK 242
216 YANAYSFWLATIERSEMFKKPTSTLKAGELRTHVSRQVCMRR 259

RESULT 11
US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2
Query Match 70.1%; Score 939; DB 4; Length 260;
Best Local Similarity 71.9%; Pred. No. 3.9e-96;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTAISCCEGTVPVLYSGSFVQGNQRAHQDGLTGLSCLORETTPPFLF 78
DB 36 GFLVTRHSQTTIDDCQPSGKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTMPFLF 95
QY 79 CNVNDVCNFSRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTVCEGPAIAIAVHSQT 138
DB 96 CNINNVNCFASRNDYSYWLSTPEPMPMSMAPITGENIRPFISCAVCEAPAMVAVHSQT 155
QY 139 TDIPCPHGWISLWKGSFIMFTSAGSEGTQOALASPGSCLEBFRSAPFIECHGRGTCNY 198
DB 156 IQIPPCSGWSSLWIGYSFVWHTSAGAGSGQALASPGSCLEBFRSAPFIECHGRGTCNY 215
QY 199 YSNSYSFWLASLNPERMFRKPIPTSTVKAGELKLIISRCQVCMKK 242
DB 216 YANAYSFWLATIERSEMFKKPTSTLKAGELRTHVSRQVCMRR 259

RESULT 12
US-09-589-987-2
; Sequence 2, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-2
Query Match 70.1%; Score 939; DB 4; Length 260;
Best Local Similarity 71.9%; Pred. No. 3.9e-96;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTAISCCEGTVPVLYSGSFVQGNQRAHQDGLTGLSCLORETTPPFLF 78
DB 36 GFLVTRHSQTTIDDCQPSGKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTMPFLF 95
QY 79 CNVNDVCNFSRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTVCEGPAIAIAVHSQT 138
DB 96 CNINNVNCFASRNDYSYWLSTPEPMPMSMAPITGENIRPFISCAVCEAPAMVAVHSQT 155
QY 139 TDIPCPHGWISLWKGSFIMFTSAGSEGTQOALASPGSCLEBFRSAPFIECHGRGTCNY 198
DB 156 IQIPPCSGWSSLWIGYSFVWHTSAGAGSGQALASPGSCLEBFRSAPFIECHGRGTCNY 215
QY 199 YSNSYSFWLASLNPERMFRKPIPTSTVKAGELKLIISRCQVCMKK 242
DB 216 YANAYSFWLATIERSEMFKKPTSTLKAGELRTHVSRQVCMRR 259

US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4
Query Match 89.0%; Score 1192; DB 3; Length 218;
Best Local Similarity 99.5%; Pred. No. 2.3e-124;
Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 QTTAISCCEGTVPVLYSGSFVQGNQRAHQDGLTGLSCLORETTPPFLFCNVNDVCN 86
DB 1 QTTAISCCEGTVPVLYSGSFVQGNQRAHQDGLTGLSCLORETTPPFLFCNVNDVCN 60
QY 87 FASRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTVCEGPAIAIAVHSQTTIDPCPH 146
DB 61 FASRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTVCEGPAIAIAVHSQTTIDPCPH 120
QY 147 GWISLWKGSFIMFTSAGSEGTQOALASPGSCLEBFRSAPFIECHGRGTCNYSNSYSFW 206
DB 121 GWISLWKGSFIMFTSAGSEGTQOALASPGSCLEBFRSAPFIECHGRGTCNYSNSYSFW 180
QY 207 LASLNPERMFRKPIPTSTVKAGELKLIISRCQVCMKKH 244
DB 181 LASLNPERMFRKPIPTSTVKAGELKLIISRCQVCMKKRH 218

RESULT 10
US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2
Query Match 70.1%; Score 939; DB 4; Length 260;
Best Local Similarity 71.9%; Pred. No. 3.9e-96;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTAISCCEGTVPVLYSGSFVQGNQRAHQDGLTGLSCLORETTPPFLF 78
DB 36 GFLVTRHSQTTIDDCQPSGKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTMPFLF 95
QY 79 CNVNDVCNFSRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTVCEGPAIAIAVHSQT 138
DB 96 CNINNVNCFASRNDYSYWLSTPEPMPMSMAPITGENIRPFISCAVCEAPAMVAVHSQT 155
QY 139 TDIPCPHGWISLWKGSFIMFTSAGSEGTQOALASPGSCLEBFRSAPFIECHGRGTCNY 198
DB 156 IQIPPCSGWSSLWIGYSFVWHTSAGAGSGQALASPGSCLEBFRSAPFIECHGRGTCNY 215
QY 199 YSNSYSFWLASLNPERMFRKPIPTSTVKAGELKLIISRCQVCMKK 242
DB 216 YANAYSFWLATIERSEMFKKPTSTLKAGELRTHVSRQVCMRR 259
```

QY 79 CNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCEGPAIAIAVHSOT 138
Db 96 CNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCEGPAIAIAVHSOT 155
QY 139 TDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFPFLECHGRGTCNY 198
Db 156 IQIPPCFSGWSSLWIGYFVWHTSAGAGSGQALASPGSCLEEFRAFPFLECHGRGTCNY 215
QY 199 YNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 242
Db 216 YANAYFWLATIERSEMFKPTSTLKGALRTHVSRQVCMKR 259

RESULT 13

US-09-589-927-10
; Sequence 10, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-10

Query Match 70.0%; Score 938.5; DB 4; Length 264;
Best Local Similarity 69.1%; Pred. No. 4.5e-96;
Matches 163; Conservative 34; Mismatches 38; Indels 1; Gaps 1;
QY 8 DSGSPATWT-TRGFVTRHSOTTAIPSCPEGTVPVLYSGFSLFVQGNORAHGQDLGTLS 66
Db 28 DKGPPTSSVAHGFLITRHSOTTDAPQCPQGTQVYEGFSLLYVQGNKRAHGQDLGTLS 87
QY 67 CLORFTTMEFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCE 126
Db 88 CLRRFTTMEFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCE 147
QY 127 GPAIAIAVHSOTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFP 186
Db 148 APAVAVIAVHSOTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFP 207
QY 187 FLECHGRGTCNYNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 242
Db 208 FIECHGRGTCNYNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 263

RESULT 14

US-09-277-665-10
; Sequence 10, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-10

Query Match 70.0%; Score 938.5; DB 4; Length 264;
Best Local Similarity 69.1%; Pred. No. 4.5e-96;
Matches 163; Conservative 34; Mismatches 38; Indels 1; Gaps 1;
QY 8 DSGSPATWT-TRGFVTRHSOTTAIPSCPEGTVPVLYSGFSLFVQGNORAHGQDLGTLS 66
Db 28 DKGPPTSSVAHGFLITRHSOTTDAPQCPQGTQVYEGFSLLYVQGNKRAHGQDLGTLS 87
QY 67 CLORFTTMEFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCE 126
Db 88 CLRRFTTMEFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCE 147
QY 127 GPAIAIAVHSOTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFP 186
Db 148 APAVAVIAVHSOTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFP 207
QY 187 FLECHGRGTCNYNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 242
Db 208 FIECHGRGTCNYNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 263

RESULT 15

US-09-589-987-10
; Sequence 10, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-10

Query Match 70.0%; Score 938.5; DB 4; Length 264;
Best Local Similarity 69.1%; Pred. No. 4.5e-96;
Matches 163; Conservative 34; Mismatches 38; Indels 1; Gaps 1;
QY 8 DSGSPATWT-TRGFVTRHSOTTAIPSCPEGTVPVLYSGFSLFVQGNORAHGQDLGTLS 66
Db 28 DKGPPTSSVAHGFLITRHSOTTDAPQCPQGTQVYEGFSLLYVQGNKRAHGQDLGTLS 87
QY 67 CLORFTTMEFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCE 126
Db 88 CLRRFTTMEFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALEPIVSRCTVCE 147
QY 127 GPAIAIAVHSOTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFP 186
Db 148 APAVAVIAVHSOTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRAFP 207
QY 187 FLECHGRGTCNYNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 242
Db 208 FIECHGRGTCNYNSYSFWLASLNPFRKPISTVKAGELEKIISRCQVCMKK 263

Search completed: April 5, 2004, 07:07:23
Job time : 58.3075 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.14528 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221B-37

Perfect score: 147

Sequence: 1 TMAPFLFCNVNDVCFNFSRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 147 | 100.0 | 220 | 2 B49736 | collagen alpha 3(I) |
| 2 | 147 | 100.0 | 1670 | 1 CGHU3B | collagen alpha 3(I) |
| 3 | 146 | 99.3 | 471 | 2 A39024 | collagen alpha 3(I) |
| 4 | 141 | 95.9 | 161 | 2 S49488 | collagen alpha 3(I) |
| 5 | 141 | 95.9 | 246 | 2 I48302 | collagen alpha 3(I) |
| 6 | 141 | 95.9 | 258 | 2 B12228 | collagen alpha 1(I) |
| 7 | 141 | 95.9 | 1669 | 1 CGHU4B | collagen alpha 1(I) |
| 8 | 141 | 95.9 | 1669 | 1 CGMS4B | collagen alpha 1(I) |
| 9 | 139 | 94.6 | 253 | 2 I48304 | collagen alpha 5(I) |
| 10 | 139 | 94.6 | 754 | 2 A5267 | collagen alpha 5(I) |
| 11 | 139 | 94.6 | 1691 | 1 S22917 | collagen alpha 5(I) |
| 12 | 129 | 87.8 | 1752 | 2 A45407 | collagen alpha 3(I) |
| 13 | 125 | 85.0 | 1747 | 2 A4121 | collagen alpha-4 c |
| 14 | 125 | 85.0 | 1763 | 2 S16366 | collagen alpha 2(I) |
| 15 | 120 | 81.6 | 1758 | 2 T93350 | hypothetical prote |
| 16 | 120 | 81.6 | 1759 | 2 T93351 | collagen alpha 2(I) |
| 17 | 117 | 79.6 | 1744 | 2 S40991 | collagen alpha 1(I) |
| 18 | 116 | 78.9 | 261 | 2 A34476 | collagen alpha 2(I) |
| 19 | 115 | 78.2 | 1691 | 1 CGHU6B | collagen alpha 6(I) |
| 20 | 112 | 76.2 | 1775 | 2 A12228 | collagen alpha 2(I) |
| 21 | 112 | 76.2 | 1707 | 2 A33526 | collagen alpha 2(I) |
| 22 | 112 | 76.2 | 1712 | 1 CGHU2B | collagen alpha 2(I) |
| 23 | 103 | 70.1 | 312 | 2 I48303 | collagen alpha 4(I) |
| 24 | 103 | 70.1 | 623 | 2 A45137 | collagen alpha 4(I) |
| 25 | 103 | 70.1 | 1690 | 1 CGHU1B | collagen alpha 4(I) |
| 26 | 102 | 69.4 | 453 | 2 S18804 | collagen alpha 4(I) |
| 27 | 93 | 63.3 | 1775 | 2 A1893 | collagen alpha 1(I) |
| 28 | 87 | 59.2 | 1761 | 2 T13990 | collagen type IV a |
| 29 | 73 | 49.7 | 79 | 2 C43928 | probable collagen |

30 59.5 40.5 58 2 B43928
31 52.5 35.7 886 2 T39081
32 50.5 34.4 257 2 H75419
33 50.5 34.4 683 1 S46492
34 49.5 33.7 419 2 S41607
35 49 33.3 397 2 S54018
36 49 33.3 537 2 F95993
37 48 32.7 98 1 F2NTK
38 48 32.7 309 2 AC1011
39 48 32.7 380 2 S66728
40 47.5 32.3 205 2 T31046
41 47.5 32.3 676 2 F85107
42 47.5 32.3 705 2 T04052
43 47.5 32.3 1474 2 T20488
44 47 32.0 178 2 F83898
45 47 32.0 811 2 T39336

ALIGNMENTS

RESULT 1

B49736
collagen alpha 3(IV) chain, medium splice form - human (fragment)
N;Contains: collagen alpha 3(IV) chain, splice form GP-V
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text_change 17-Mar-2000
C;Accession: B49736; D49736; S69111
R;Peng, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.
A;Reference number: A49736; MUID:94124597; PMID:8294492
A;Accession: B49736
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 169-220 <FEN1>
A;Accession: D49736
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 22-220 <FEN2>
A;Cross-references: GB:U02519; NID:9409106; PIDN:AAA18942.1; PID:9409107
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wi
Eur. J. Biochem. 229, 754-760, 1995
A;Title: Characterization and expression of multiple alternatively spliced transcripts
utoc antigen and one of its alternative forms.
A;Reference number: S69111; MUID:95278230; PMID:7758473
A;Accession: S69111
A;Molecule type: mRNA
A;Residues: 1-45,169-204,'L',206-220 <PEN>
C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.
C;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-2q37
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace
F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pred
F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status
F;22-220/Domains: carboxyl-terminal nonhelical, NC1 <NC1>
F;34-134/Domains: collagen IV carboxyl-terminal repeat <CT1>

Query Match 100.0%; Score 147; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFLFCNVNDVCFNFSRNDYSYWL 25

DB 82 TMAPFLFCNVNDVCFNFSRNDYSYWL 106

RESULT 2

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human
 N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
 C/Species: Homo sapiens (man)
 C/Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
 C/Accession: A54763; A43928; A44043; A45971; A39786
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.
 J. Biol. Chem. 269, 23013-23017, 1994
 A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
 A/Reference number: A54763; MUID:94364994; PMID:8083201
 A/Accession: A54763
 A/Molecule type: mRNA
 A/Residues: 1-1670 <MAR>
 A/Cross-references: GB:X80031; NID:G577563; PID:G577564
 A/Experimental source: kidney
 R/Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Fovey, S.; Rees, A.; Pusey, C.D.
 J. Clin. Invest. 89, 592-601, 1992
 A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha
 A/Reference number: A43928; MUID:92147878; PMID:1737849
 A/Accession: A43928
 A/Molecule type: mRNA
 A/Residues: 1331-1524, 'I', 1526-1670 <TUR>
 A/Cross-references: GB:M81373
 A/Experimental source: kidney
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 267, 19780-19784, 1992
 A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
 A/Reference number: A44043; MUID:93015826; PMID:1400291
 A/Accession: A44043
 A/Molecule type: DNA; mRNA
 A/Residues: 1386-1670 <QUI>
 A/Cross-references: GB:M92993; NID:G177895; PIDN:AA21610.1; PID:G177896
 A/Note: sequence extracted from NCBI backbone (NCBI:115597)
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 269, 17358, 1994
 A/Reference number: A44738; MUID:94274734; PMID:8006044
 A/Contents: annotation; erratum; correction to intronic sequence in A44043
 R/Bernal, D.; Quinones, S.; Saus, J.
 J. Biol. Chem. 268, 12090-12094, 1993
 A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A/Reference number: A45971; MUID:93280184; PMID:8505332
 A/Accession: A45971
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1427-1444 <BER>
 A/Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly ident
 R/Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reeders, S.T.
 Am. J. Hum. Genet. 49, 545-554, 1991
 A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
 A/Reference number: A39786; MUID:91353570; PMID:1882840
 A/Accession: A39786
 A/Molecule type: mRNA
 A/Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A/Cross-references: GB:S55790; NID:G234418; PIDN:AA819637.1; PID:G234419
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
 C/Genetics:
 A/Gene: GDB:COL4A3
 A/Cross-references: GDB:128351; OMIM:120070
 A/Map position: 2q36-2q37
 A/Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1595/3; 1643/2 #status incomplete
 A/Note: The alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
 er associations in the interrupted helical domain (with disulfide and desmosine cross-l
 C/Function:
 A/Description: minor structural component of extracellular basement membrane in kidney g
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 F;1-28/Domain: signal sequence #status predicted <SIG>
 F;19-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
 F;29-42/Domain: amino-terminal nonhelical, NH1 <NHI>

F;43-1438/Region: interrupted helical
 F;791-793/Region: cell attachment (R-G-D) motif
 F;996-998/Region: cell attachment (R-G-D) motif
 F;1154-1156/Region: cell attachment (R-G-D) motif
 F;1306-1308/Region: cell attachment (R-G-D) motif
 F;1345-1347/Region: cell attachment (R-G-D) motif
 F;1432-1434/Region: cell attachment (R-G-D) motif
 F;1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F;1,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status pred
 F;253/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F;1505-1511,1616-1622/Disulfide bonds: #status predicted
 F;1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
 Query Match 100.0%; Score 147; DB 1; Length 1670;
 Best Local Similarity 100.0%; Pred. No. 1.2e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TMPFLFCNVNVCNFSRNDYSYWL 25
 |||||
 DB 1499 TMPFLFCNVNVCNFSRNDYSYWL 1523
 |||||
 RESULT 3
 A39024
 collagen alpha 3(IV) chain - bovine (fragment)
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
 C/Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
 R/Morrison, K.E.; Germino, G.G.; Reeders, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A/Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
 A/Reference number: A39024; MUID:91093146; PMID:1985905
 A/Accession: A39024
 A/Molecule type: mRNA
 A/Residues: 1-471 <MOR>
 A/Cross-references: EMBL:M63139; NID:G162886; PIDN:AAA62708.1; PID:G162887
 R/Bukowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A/Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A/Reference number: S18432; MUID:87222419; PMID:2438283
 A/Accession: S20672
 A/Molecule type: protein
 A/Residues: 227-228, 'X', 230-244 <BUT>
 R/Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A/Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
 A/Reference number: S17802; MUID:88330844; PMID:3417661
 A/Accession: S17802
 A/Molecule type: protein
 A/Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R/Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 265, 5466-5469, 1990
 A/Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of typ
 A/Reference number: A35167; MUID:90202779; PMID:2318822
 A/Accession: A35167
 A/Molecule type: protein
 A/Residues: 236-258 <GUN>
 R/Gunwar, S.; Bailester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; No
 J. Biol. Chem. 266, 15318-15324, 1991
 A/Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol
 A/Reference number: A39419; MUID:91332055; PMID:1669555
 A/Accession: C39419
 A/Molecule type: protein
 A/Residues: 236-255 <GU2>
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;
 F;1-238/Domain: collagenous (fragment) #status predicted <COL>
 F;239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
 F;239-353/Domain: repeat NC1 #status predicted <NC11>
 F;354-471/Domain: repeat NC1 #status predicted <NC12>

R;Soininen, R.; Haka-Risk
FEBS Lett. 225, 188-194,

FEBS Lett. 225, 188-194, 1987

A:Title: Complete primary structure of the alpha-1(I)-chain of human basement membrane (ty
A:Reference number: S00207; MUID:8803584; PMID:3691802
A:Accession: S00207
A:Molecule type: mRNA
A:Residues: 244-530 <SOL3>
A:Cross-references: EMBL:Y00706; MID:g29548; PIDN:CAA68698.1; PID:g29549
R:Esle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
A:Reference number: S39614; MUID:94038963; PMID:8223498
A:Accession: S39614
A:Molecule type: protein
A:Residues: 371-554 <EBL>
R:Babel, W.; Glanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid se
A:Reference number: A02863; MUID:85003629; PMID:6434307
A:Accession: A02863
A:Molecule type: protein
A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-
R:Experimental source: placenta
R:Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A:Title: Pepsin fragments of human placental basement-membrane collagens showing interr
A:Reference number: S16908; MUID:82005835; PMID:6792033
A:Accession: A58517
A:Molecule type: protein
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fieczek, P.P.
Biochemistry 22, 4940-4948, 1983
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (b
A:Reference number: S16910; MUID:84053346; PMID:6416291
A:Accession: S16910
A:Molecule type: protein
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-
R:Experimental source: placenta
R:Phillips, T.; Trygvaeson, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F
J. Biol. Chem. 260, 7681-7687, 1985
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
A:Reference number: S01466; MUID:85207819; PMID:2581959
A:Accession: S01466
A:Molecule type: mRNA
A:Residues: 1256-1669 <PTH>
A:Cross-references: EMBL:M10940; NID:g180421; PIDN:AAAS2006.1; PID:g180424
R:Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
A:Reference number: S16879; MUID:85216555; PMID:2582422
A:Accession: S16879
A:Molecule type: mRNA
A:Residues: 1259-1669 <BRI>
A:Cross-references: EMBL:M1315; NID:g180817; PIDN:AAAS2042.1; PID:g180818
R:Oberbauer, I.; Laurent, M.; Schwaz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
Eur. J. Biochem. 147, 217-224, 1985
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A:Reference number: A02864; MUID:85127033; PMID:2578961
A:Accession: S19091
A:Molecule type: protein
A:Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',
R:Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
A:Reference number: S02550; MUID:89005112; PMID:2844531
A:Contents: annotation; disulfide bonds
C:Genetics:
A:Gene: GDB:COL4A1
A:Cross-references: GDB:119791; OMIM:120130
A:Map position: 13q34-13q34
A:Intron: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/
A:Exon: 1/1; 31/3; 782/1; 820/1; 876/1; 908/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
C:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(
C:Associations: among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
C:Associations: in the interrupted helical domain (with disulfide and desmosine cr

C:Function:
A:Description: structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication;
F:1-36/Domain: signal sequence #status predicted <SIG>
F:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F:29-162/Domain: amino-terminal nonhelical, 78 <78>
F:163-1440/Domain: interrupted helical <COL>
F:414-452/Region: integrin binding #status experimental
F:597-599/Region: cell attachment (R-G-D) motif
F:917-919/Region: cell attachment (R-G-D) motif
F:968-970/Region: cell attachment (R-G-D) motif
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:27/Modified site: blocked amino end (Ala) (in mature form) #status experimental
F:31.36.39.41.125.434.467.470/Disulfide bonds: interchain #status predicted
F:45.48.78.90.129.156.172.217.228.231.277.295.298.322.343.361.460.463.497.527.540.543.5.
1081.1084.1099.1117.1132.1150.1165.1182.1185.1188.1206.1235.1265.1283.1304.1319.1328.13.
F:45.48.78.90.129.156.217.228.231.277.295.298.322.343.361.460.463.497.527.543.573.582.6.
99.1117.1132.1150.1165.1182.1185.1188.1206.1235.1265.1283.1304.1319.1328.1340.1356.1371.
F:54.63.75.84.87.96.102.105.108.111.117.120.123.136.141.147.150.153.159.167.178.181.184.
F:419.422.425.439.445.448.451.479.485.491.494.503.512.518.524.530.546.549.552.555.561.56.
9.745.748.751.754.763/Modified site: 4-hydroxyproline (Pro) #status experimental
F:126/Binding site: carboxylate (Asn) (covalent) #status experimental
F:129/Modified site: allysine (Lys) #status predicted
F:172.540.947/Modified site: 5-hydroxylysine (Lys) #status atypical
F:272.645.839/Modified site: 4-hydroxyproline (Pro) #status atypical
F:446.475.784.787.790.796.804.810.816.822.834.860.863.869.872.875.887.890.893.899.9.
F:1129.1138.1141.1159.1171.1176.1179.1194.1200.1203.1215.1224.1227.1244.1247.1250.1256.
431.1437/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1120.1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F:1120.1268/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental
F:1215.1424/Modified site: 3-hydroxyproline (Pro) #status absent
F:1460-1548.1493-1551/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1505-1511.1616-1622/Disulfide bonds: #status predicted
F:1570-1662.1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
Query Match 95.9%; Score 141; DB 1; Length 1669;
Best Local Similarity 92.0%; Pred No. 8.3e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMAPFLFCNVNVCNFCASNDYSYL 25
DB 1499 TMAPFLFCNVNVCNFCASNDYSYL 1523
RESULT 8
CGMS4B
collagen alpha 1(IV) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000
C:Accession: S01452; S01454; A28065; A2864; A25636; A29301; S19079; A32003; A31766; S1.
R:Mutukumar, G.; Blumberg, B.; Kurkinen, M.
J. Biol. Chem. 264, 6310-6317, 1989
A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Dif
A:Reference number: A33525; MUID:89197932; PMID:2703490
A:Accession: A33525
A:Molecule type: mRNA
A:Residues: 1-1669 <MUT>
A:Cross-references: ENBL:J04694; NID:g556296; PIDN:AAAS0292.1; PID:g556297
R:Wood, L.; Theriault, N.; Vogeli, G.
FEBS Lett. 227, 5-8, 1988
A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the a.
A:Reference number: S01454; MUID:88112221; PMID:3338568
A:Accession: S01454
A:Molecule type: mRNA
A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-7.
A:Cross-references: EMBL:X06777
R:Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.
J. Biol. Chem. 263, 8706-8709, 1988

A:Reference number: A25991; MUID:82186723; PMID:6804236

A:Accession: A25991

A:Molecule type: protein

A:Residues: 940-946,'X','948-949,'X','951-955,'X','957-964,'X','966-991,'X','993-1003,'X','1010-1011,'X','1063-1065,'X','1067-1080,'X','1082-1083,'X','1085-1106,'X','1108-1115,'DE','1118-1119

A:Accession: B25991

A:Molecule type: protein

A:Residues: 1173-1181,'X','1183-1184,'X','1186-1187,'X','1189-1205,'O','1207,'XE','1210-1234

3,'SP','1266,'IT','1269,'SK','1272,'DM','1275,'L','1277-1282','1316-1318,'X','1320-1327,'X','1332

R:Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.

Eur. J. Biochem. 139, 401-410, 1984

A:Title: Subunit structure and assembly of the globular domain of basement-membrane col

A:Reference number: S17801; MUID:84132058; PMID:6698021

A:Accession: S17801

A:Molecule type: protein

A:Residues: 1435-1443 <WEB>

C:Genetics:

A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3

A:Note: the list of introns may be incomplete

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; cell binding; coiled coil, duplication; extracellular ma

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>

F:28-162/Domain: 7S <SDS>

F:163-1440/Domain: collagenous, triple helix <COL>

F:597-599/Region: cell attachment (R-G-D) motif

F:781-783/Region: cell attachment (R-G-D) motif

F:917-919/Region: cell attachment (R-G-D) motif

F:968-970/Region: cell attachment (R-G-D) motif

F:1441-1669/Domain: carboxyl-terminal nonhelical, NCI <NCL>

F:1441-1552/Region: duplication

F:1553-1669/Region: duplication

F:31,36,39,41,434,467,470/Disulfide bonds: interchain #status predicted

F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

E:911,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,1090,

92,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxypr

F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental

F:1304/Modified site: 5-hydroxylysine (lys) #status experimental

F:1505-1511,1616-1622/Disulfide bonds: #status predicted

Query Match 95.9%; Score 141; DB 1; Length 1669;

Best Local Similarity 92.0%; Pred. No. 8,3e-12;

Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCNFSARNDSYWL 25

Db 1499 TMPFLFCNINNVNCFASRNDSYWL 1523

RESULT 9

148304

collagen alpha 5(IV) chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999

C:Accession: I48304; S47280

R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48304

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-253 <RES>

A:Cross-references: EMBL:Z55168, NID:9535201; PIDN:CAA84531.1; PID:9535202

C:Superfamily: collagen alpha 1(IV) chain

Query Match 94.6%; Score 139; DB 2; Length 253;

Best Local Similarity 89.0%; Pred. No. 2,8e-12;

Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCNFSARNDSYWL 25

Db 83 TMPFMFCNINNVNCFASRNDSYWL 107

[illegible]

Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
|||||:|:|:|:|:|:|
Db 1587 TMPFLFCDVNNVCNYASRNDKSYWL 1611

RESULT 15

T29350

Hypochemical protein F01G12.5a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T29350

R:Wu, X.; Le, T.T.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F01G12.

A:Reference number: Z20611

A:Accession: T29350

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1758 <WUX>

A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GNC00028; CESP:F01G12.5a

A:Experimental source: strain Bristol N2; clone F01G12

C:Genetics:

A:Gene: CESP:F01G12.5a

A:Map position: X

A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3

C:Superfamily: collagen alpha 1(IV) chain

Query Match

81.6%; Score 120; DB 2; Length 1758;

Best Local Similarity 80.0%; Fred. NO. 8.1e-09;

Matches 20; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
|||||:|:|:|:|:|:|
Db 1585 TMPFLFCDVNNVCNYASRNDKSYWL 1609

Search completed: April 5, 2004, 07:05:35
Job time : 6.14528 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.1477 Seconds
(without alignments)
413.557 Million cell updates/sec

Title: US-10-032-221B-37
Perfect score: 147
Sequence: 1 TMAPLFCNVNDVCFASRNDYSYWL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|--------------|---------------------|
| 1 | 147 | 100.0 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 2 | 146 | 99.3 | 471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 3 | 141 | 95.9 | 1669 | 1 CA14 HUMAN | P02462 homo sapien |
| 4 | 141 | 95.9 | 1669 | 1 CA14 MOUSE | P02463 mus musculus |
| 5 | 139 | 94.6 | 754 | 1 CA54 CANFA | Q28247 canis famil |
| 6 | 139 | 94.6 | 1685 | 1 CA54 HUMAN | P29400 homo sapien |
| 7 | 125 | 85.0 | 1763 | 1 CA24 ASCSU | P27393 ascaris suu |
| 8 | 117 | 79.6 | 1758 | 1 CA14 CAEEL | P17139 caenorhabdi |
| 9 | 116 | 78.9 | 1758 | 1 CA24 CAEEL | P17140 caenorhabdi |
| 10 | 115 | 78.2 | 1691 | 1 CA64 HUMAN | Q14031 homo sapien |
| 11 | 112 | 76.2 | 1707 | 1 CA24 MOUSE | P08122 mus musculus |
| 12 | 112 | 76.2 | 1712 | 1 CA24 HUMAN | P08572 homo sapien |
| 13 | 103 | 70.1 | 623 | 1 CA44 RABIT | P55787 oryctolagus |
| 14 | 103 | 70.1 | 1690 | 1 CA44 HUMAN | P53420 homo sapien |
| 15 | 102 | 69.4 | 453 | 1 CA44 BOVIN | Q29442 bos taurus |
| 16 | 93 | 63.3 | 1775 | 1 CA14 DROME | P08120 drosophila |
| 17 | 51 | 34.7 | 333 | 1 AMR1 HUMAN | Q9Y4X0 homo sapien |
| 18 | 51 | 34.7 | 344 | 1 AMR1 MOUSE | Q9Y4X5 mus musculus |
| 19 | 50.5 | 34.4 | 663 | 1 VM02 CHICK | Q90611 gallus gall |
| 20 | 49 | 33.3 | 397 | 1 VM07 YEAST | Q04359 saccharomyc |
| 21 | 48 | 32.7 | 308 | 1 META_SALTI | Q821W1 salmonella |
| 22 | 48 | 32.7 | 308 | 1 META_SALTY | P37413 salmonella |
| 23 | 48 | 32.7 | 380 | 1 NTG2 YEAST | Q08214 saccharomyc |
| 24 | 47 | 32.0 | 186 | 1 RYGL HUMAN | Q9BW83 homo sapien |
| 25 | 46.5 | 31.6 | 379 | 1 MARG PORGI | Q7MAW5 porphyron |
| 26 | 46 | 31.0 | 356 | 1 CMO_EYRFU | Q51741 pyrococcus |
| 27 | 45.5 | 31.0 | 1743 | 1 TACG_DICDI | Q23868 dictyosteli |
| 28 | 45 | 30.6 | 1095 | 1 C2S5_SACKL | Q00704 autographa |
| 29 | 44.5 | 30.3 | 704 | 1 OE66_NPVAC | Q13145 homo sapien |
| 30 | 44 | 29.9 | 260 | 1 NMA HUMAN | Q17917 caenorhabdi |
| 31 | 44 | 29.9 | 384 | 1 Y092 CAEEL | Q9ZE55 rickettsia |
| 32 | 43.5 | 29.6 | 334 | 1 Y092_RICPR | P20061 homo sapien |
| 33 | 43.5 | 29.6 | 433 | 1 TC01_HUMAN | |

34 43.5 29.6 662 1 MM02 MOUSE P33434 mus musculus
35 43.5 29.6 662 1 MM02 RAT P33436 rattus norv
36 43 29.3 128 1 CD59 HUMAN P13987 h cd59 glyc
37 43 29.3 343 1 Z183 HUMAN O15541 homo sapien
38 43 29.3 385 1 CHEB_BORBU Q45047 borellia bu
39 43 29.3 395 1 NH10_CAEL P41999 caenorhabdi
40 43 29.3 464 1 SYR2_COXBU Q83616 coxiella bu
41 43 29.3 849 1 SRK6_BRAOL Q09092 brassica ol
42 42.5 28.9 361 1 ALR_CORGL Q8REU9 corynebacte
43 42.5 28.9 407 1 NKIR_MOUSE P25103 homo sapien
44 42.5 28.9 407 1 NKIR_MOUSE P30548 mus musculus
45 42.5 28.9 407 1 NKIR_RAT P14600 rattus norv

ALIGNMENTS

RESULT 1
CA34_HUMAN STANDARD; PRT; 1670 AA.
ID CA34_HUMAN STANDARD; PRT; 1670 AA.
AC Q01955; Q9BOT2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Eutelestostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN R1
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Readers S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. Biol. Chem. 269:23013-23017(1994).
[2]
RN R2
RP REVISIONS.
RA Leinonen A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN R3
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064696; PubMed=1134955;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RT Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
[4]
RN R4
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015836; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:19780-19784(1992).
[5]
RN R5
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1892840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Readers S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
[6]
RN R6
RP SEQUENCE OF 1331-1670 FROM N.A.
RX TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;

RP SEQUENCE OF 1-943 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Brazel D., Oberbauer I., Diezinger H., Babel W., Glanville R.W.,
RA Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
RT basement membrane collagen (type IV) reveals 21 non-triplet
RT interruptions located within the collagenous domain.";
RL Eur. J. Biochem. 168:529-536(1987).
RN [4]
RP SEQUENCE OF 28-243.
RX MEDLINE=86004708; PubMed=4043082;
RA Glanville R.W., Qian R.O., Siebold B., Risteli J., Kuehn K.;
RT "Amino acid sequence of the N-terminal aggregation and cross-linking
RT region (7S domain) of the alpha 1 (IV) chain of human basement
RT membrane collagen.";
RL Eur. J. Biochem. 152:213-219(1985).
RN [5]
RP SEQUENCE OF 534-1447.
RX MEDLINE=85003629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RT "Structure of human-basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1(IV) chain.";
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85207819; PubMed=2581969;
RA Phlajajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA Cheung W.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha 1(IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxy-terminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Lysines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
unit (G-X-Y) are hydroxylated in some or all of the chains.
unit (G-X-Y) are hydroxylated in some or all of the chains.
these, located in the NC1 domain, are conserved in all known type
IV collagens.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch)
EMBL; M26576; AAA53098.1; JOINED.
EMBL; J04217; AAA53098.1; JOINED.
EMBL; M26550; AAA53098.1; JOINED.
EMBL; M26540; AAA53098.1; JOINED.
EMBL; M26542; AAA53098.1; JOINED.
EMBL; M26543; AAA53098.1; JOINED.
EMBL; M26544; AAA53098.1; JOINED.
EMBL; M26545; AAA53098.1; JOINED.
EMBL; M26546; AAA53098.1; JOINED.
EMBL; M26547; AAA53098.1; JOINED.
EMBL; M26537; AAA53098.1; JOINED.
EMBL; M26538; AAA53098.1; JOINED.
EMBL; M26548; AAA53098.1; JOINED.
EMBL; M26549; AAA53098.1; JOINED.
EMBL; M26551; AAA53098.1; JOINED.
EMBL; M26552; AAA53098.1; JOINED.
EMBL; M26553; AAA53098.1; JOINED.
EMBL; M26554; AAA53098.1; JOINED.
EMBL; M26555; AAA53098.1; JOINED.
EMBL; M26556; AAA53098.1; JOINED.
EMBL; M26557; AAA53098.1; JOINED.
EMBL; M26558; AAA53098.1; JOINED.
EMBL; M26559; AAA53098.1; JOINED.
EMBL; M26560; AAA53098.1; JOINED.
EMBL; M26561; AAA53098.1; JOINED.
EMBL; M26562; AAA53098.1; JOINED.
EMBL; M26563; AAA53098.1; JOINED.
EMBL; M26564; AAA53098.1; JOINED.
EMBL; M26565; AAA53098.1; JOINED.
EMBL; M26566; AAA53098.1; JOINED.
EMBL; M26567; AAA53098.1; JOINED.
EMBL; M26568; AAA53098.1; JOINED.
EMBL; M26569; AAA53098.1; JOINED.
EMBL; M26570; AAA53098.1; JOINED.
EMBL; M26571; AAA53098.1; JOINED.
EMBL; M26572; AAA53098.1; JOINED.
EMBL; M26573; AAA53098.1; JOINED.
EMBL; M26574; AAA53098.1; JOINED.
EMBL; M26575; AAA53098.1; JOINED.
EMBL; Y00706; CAA8698.1; -
EMBL; X05561; CAA29075.1; -
EMBL; M10940; AAA52006.1; -
EMBL; M11315; AAA52042.1; -
PIR; S16876; CGHUB.
Genew; HGNC:2202; COL4A1.
MIM; 120130; -
InterPro; IPR008161; Clg_helix.
InterPro; IPR008160; Collagen.
InterPro; IPR001442; Procollagnc4_C.
Pfam; PF01413; C4; 2.
Pfam; PF01391; Collagen; 24.
ProDom; PD000007; Clg_helix; 6.
ProDom; PD003923; Procollagnc4; 1.
SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NON-HELICAL REGION (NC1).
 FT CARBOHYD 126 136 N-LINKED (GLCNAC. . .).
 FT DISULFID 1460 1551 OR 1548.
 FT DISULFID 1493 1548 OR 1551.
 FT DISULFID 1505 1511 OR 1662.
 FT DISULFID 1570 1665 OR 1665.
 FT DISULFID 1604 1662 OR 1665.
 FT DISULFID 1616 1622 OR 1665.
 FT CONFLICT 237 238 SG -> KE (IN REF. 4).
 FT CONFLICT 241 241 G -> K (IN REF. 4).
 FT CONFLICT 319 319 Q -> A (IN REF. 3).
 FT CONFLICT 719 719 N -> D (IN REF. 5).
 FT CONFLICT 837 837 D -> Y (IN REF. 5).
 FT CONFLICT 842 842 K -> P (IN REF. 5).
 FT CONFLICT 896 896 V -> W (IN REF. 2).
 FT CONFLICT 904 904 E -> Q (IN REF. 5).
 FT CONFLICT 914 914 S -> K (IN REF. 5).
 FT CONFLICT 998 998 K -> K (IN REF. 5).
 FT CONFLICT 1010 1010 S -> P (IN REF. 5).
 FT CONFLICT 1012 1012 K -> K (IN REF. 5).
 FT CONFLICT 1012 1012 S -> K (IN REF. 5).
 FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
 SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFB9B8A84 CRC64;
 Query Match 95.9%; Score 141; DB 1; Length 1669;
 Best Local Similarity 92.0%; Pred. No. 4.3e-12;
 Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWPFFCNVNDVNCNFSRNDYSYWL 25
 |||||:|||||
 Db 1499 TWPFFCNVNDVNCNFSRNDYSYWL 1523
 |||||:|||||
 RESULT 4
 ID CA14 MOUSE STANDARD; PRT; 1669 AA.
 AC P02463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89197932; PubMed=2703490;
 RA Muthukumar G., Blumberg B., Kurkinen M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 collagen IV. Differential evolution of collagen IV domains.";
 RL J. Biol. Chem. 264:6310-6317(1989).
 RN [2]
 RP SEQUENCE OF 1-1154 FROM N.A.
 RX MEDLINE=88112221; PubMed=3338568;
 RA Wood L., Theriault N., Vogeli G.;
 RT "cDNA clones completing the nucleotide and derived amino acid
 sequence of the alpha 1 chain of basement membrane (type IV) collagen
 from mouse.";
 RL FEBS Lett. 227:5-8(1988).
 RN [3]
 RP SEQUENCE OF 1149-1424 FROM N.A.
 RX MEDLINE=86101886; PubMed=3755692;
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
 synthetic oligodeoxynucleotide.";
 RL Gene 43:301-304(1986).
 RN [4]
 RP SEQUENCE OF 1276-1669 FROM N.A.
 RX MEDLINE=85127033; PubMed=2578961;
 RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
 Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
 the alpha 1(IV) chain of basement membrane collagen as derived from
 complementary DNA.";
 RL Eur. J. Biochem. 147:217-224(1985).
 RN [5]
 RP SEQUENCE OF 1441-1669 FROM N.A.
 RX MEDLINE=87250460; PubMed=3597383;
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
 Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse
 alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8496-8499(1987).
 RN [6]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=86196099; PubMed=3009468;
 RA Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
 collagen genes.";
 RL J. Biol. Chem. 261:6654-6657(1986).
 RN [7]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89066738; PubMed=3198626;
 RA Kayes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
 RT "Head-to-head arrangement of murine type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277(1988).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89071759; PubMed=3200851;
 RA Burdello P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
 bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
 RN [9]
 RP SEQUENCE OF 1-129 FROM N.A.
 RX MEDLINE=88243724; PubMed=3379041;
 RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
 collagen chain and the corresponding region of the gene.";
 RL J. Biol. Chem. 263:8706-8709(1988).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 glomerular basement membrane (GBM), forming a 'chicken-wire'
 meshwork together with laminins, proteoglycans and entactin/
 nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 alpha 6(IV), each of which can form a triple helix structure with
 2 other chains to generate type IV collagen network.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the G-
 X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J03758; AAA37439.1; -.
 CC EMBL; M23333; AAA51625.1; -.

```
DR EMBL; J04694; A0450292.1; -
DR EMBL; X06777; CAA29946.1; -
DR EMBL; X02201; CAA26132.1; -
DR EMBL; M15832; AAA37340.1; -
DR EMBL; M14042; AAA37343.1; -
DR EMBL; M12879; AAA37343.1; -
DR EMBL; M13024; -; NOT_ANNOTATED_CDS.
DR EMBL; M13025; -; NOT_ANNOTATED_CDS.
DR EMBL; M13026; AAA37344.1; -
DR EMBL; M13027; AAA37345.1; -
DR EMBL; M13043; AAA37346.1; -
DR EMBL; J04448; AAA37437.1; -
DR PIR; A33525; CGMS4B.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1662 BY SIMILARITY.
FT CARBOHYD 126 136 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 882 982 Q -> H (IN REF. 3).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 95.9%; Score 141; DB 1; Length 1669;
Best Local Similarity 92.0%; Pred. No. 4.3e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVNCNFSRNDYSYL 25
Db 1499 TMPFLFCNVNDVNCNFSRNDYSYL 1523

RESULT 5
ID_CAS4 CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; TISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thorner P.S., Marrano P., Bauml R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
human X-linked hereditary nephritis resulting from a single base
mutation in the gene encoding the alpha 5 chain of collagen type
IV.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the G-
X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
canine X-linked hereditary nephritis (HN), a disease similar to
that in humans (also referred to as Alport syndrome) characterized
by progressive renal failure and neurosensory deafness.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U07889; AAB60258.1; -
CC PIR; A55267; A55267.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagn4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 8.
CC ProDom; PD000007; Clg_helix; 1.
CC ProDom; PD003923; ProcollagnC4; 1.
CC SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT DOMAIN 1 530 TRIPLE-HELICAL REGION.
FT DOMAIN 531 754 NONHELICAL REGION (NC1).
FT DISULFID 552 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 7 OR 754 (BY SIMILARITY).
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON_TER 754 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 94.6%; Score 139; DB 1; Length 754;
Best Local Similarity 88.0%; Pred. No. 3.7e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVNCNFSRNDYSYL 25
Db 591 TMPFLFCNVNDVNCNFSRNDYSYL 615
```

RESULT 6

CA54_HUMAN STANDARD; PRT; 1685 AA.
ID AC CA54_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337930; PubMed=2380186;
RA Pihlajaniemi T., Pohjola-Erkkonen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169451; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Myers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjola-Erkkonen E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;

RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Sheets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
RN [14]
RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;
RX ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
RN [15]
RP VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchino C.J., Jiang X., Atkin C.L., Stone E.M., Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
RN [16]
RP VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S., Turco A.B., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G., Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F., Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51 exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).

RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=597094179; PubMed=8940267;
RA Knebelmann B., Brüllat C., Forestier L., Arrondel C., Jacassier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Brocay M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
RN [18]
RP VARIANT AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobrykina V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of $\alpha 5(\text{IV})$ collagen
chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
RN [19]
RP VARIANT AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
RN [20]
RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
RN [21]
RP VARIANTS AS VAL-420; 456-PRO-PRO-458 DEL; ASP-573; ASP-624; ASP-635;
802-GLY-PRO DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
RX MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Styrikardottir U., Neumann H.,
RA Spingate J., Shows T.B., Pettersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
RN [22]
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
patients with X-linked Alport's syndrome by RT-PCR and direct
sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
RN [23]
RP VARIANT AS ARG-822.
Query Match 94.6%; Score 139; DB 1; Length 1685;
Best Local Similarity 88.0%; Pred. No. 8.4e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMBPFCNVNDVNCNFSRNDYSYWL 25
Db 1515 TMBPFCNVNDVNCNFSRNDYSYWL 1539
RESULT 7
CA24-ASCSU STANDARD; PRT; 1763 AA.
AC P27393;
DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS I AND II).
RX MEDLINE=91340768; PubMed=1714907;
RA Pettitt J., Kingston I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
and the partial structural organization of its gene.";
RL J. Biol. Chem. 266:16149-16156(1991).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I;
CC IsoId=P27393-1; Sequence=Displayed;
CC Name=II;
CC IsoId=P27393-2; Sequence=VSP_001159;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M675507; AAA18014.1; -.
CC PIR; S16366; S16366.
CC InterPro; IPR008161; Clg helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagn4_C.
CC Pfam; PF01413; C4; 2.
CC ProDom; PD000007; Clg helix; 6.
CC ProDom; PD003923; ProcollagnC4; 1.
CC SMART; SMC0111; C4; 2.
KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
KW Alternative splicing; Glycoprotein; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 27 42 7S DOMAIN.
FT DOMAIN 43 1529 TRIPLE-HELICAL REGION.
FT DOMAIN 1530 1763 NONHELICAL REGION (NC1).
FT DISULFID 1548 1637 OR 1634 (BY SIMILARITY).
FT DISULFID 1581 1634 OR 1637 (BY SIMILARITY).
FT DISULFID 1593 1599 BY SIMILARITY.
FT DISULFID 1656 1752 OR 1749 (BY SIMILARITY).
FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).
FT DISULFID 1702 1709 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 249 249 O-LINKED (XYL...) (GLYCOSAMINOGLYCAN)
FT (IN ISOFORM II) (POTENTIAL).
FT VARSPLIC 230 266 GEQPGPGQPGFVSTGAKGTITGPGAGMKGK ->
GTGPGAGPGPGPGPREFTGSGIVGRHSGDKGVK (in


```
FT isoform II).
FT /FTID=VSP_001159.
SQ SEQUENCE 1763 AA; 168526 MW; 304F52B8C06A80D CRC64;

Query Match
Best Local Similarity 85.0%; Score 125; DB 1; Length 1763;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 TMPPFLFCNVNDVCFASRNDYSYL 25
DB 1587 TMPPFLFCNVNDVCFASRNDYSYL 1611

RESULT 8
CA14_CABEL
ID CA14_CABEL STANDARD; PRT; 1758 AA.
AC P17139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR C1B-2 OR X04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsay T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Lauster N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlschlag J.,
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -1- FUNCTION: Collagen type IV is specific for basement membranes.
CC -1- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
intermolecular interactions between 7S domains and between NC1
domains.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
```


OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eye, and Kidney;
RX MEDLINE=9411779; PubMed=8125972;
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.,
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5".
RL J. Biol. Chem. 269:7520-7526 (1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Redders S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199 (1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Redders S.T., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grafham D., Lawlor S., Wilson S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=8356449;
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,
RA de Raepae A., Tryggvason K., Redders S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169 (1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D21337; BAA04809.1; JOINED.
CC EMBL; U04845; AAB19038.1; JOINED.
CC EMBL; U47004; AAB19038.1; JOINED.
CC EMBL; U46959; AAB19038.1; JOINED.
CC EMBL; U46961; AAB19038.1; JOINED.
CC EMBL; U46962; AAB19038.1; JOINED.
CC EMBL; U46963; AAB19038.1; JOINED.
CC EMBL; U46964; AAB19038.1; JOINED.
CC EMBL; U46965; AAB19038.1; JOINED.
CC EMBL; U46966; AAB19038.1; JOINED.
CC EMBL; U46967; AAB19038.1; JOINED.
CC EMBL; U46968; AAB19038.1; JOINED.
CC EMBL; U46969; AAB19038.1; JOINED.
CC EMBL; U46970; AAB19038.1; JOINED.
CC EMBL; U46971; AAB19038.1; JOINED.
CC EMBL; U46972; AAB19038.1; JOINED.
CC EMBL; U46973; AAB19038.1; JOINED.
CC EMBL; U46974; AAB19038.1; JOINED.
CC EMBL; U46975; AAB19038.1; JOINED.
CC EMBL; U46976; AAB19038.1; JOINED.
CC EMBL; U46977; AAB19038.1; JOINED.
CC EMBL; U46978; AAB19038.1; JOINED.
CC EMBL; U46979; AAB19038.1; JOINED.
CC EMBL; U46980; AAB19038.1; JOINED.
CC EMBL; U46981; AAB19038.1; JOINED.
CC EMBL; U46982; AAB19038.1; JOINED.
CC EMBL; U46983; AAB19038.1; JOINED.
CC EMBL; U46984; AAB19038.1; JOINED.
CC EMBL; U46985; AAB19038.1; JOINED.
CC EMBL; U46986; AAB19038.1; JOINED.
CC EMBL; U46987; AAB19038.1; JOINED.
CC EMBL; U46988; AAB19038.1; JOINED.
CC EMBL; U46989; AAB19038.1; JOINED.
CC EMBL; U46990; AAB19038.1; JOINED.
CC EMBL; U46991; AAB19038.1; JOINED.
CC EMBL; U46992; AAB19038.1; JOINED.
CC EMBL; U46993; AAB19038.1; JOINED.
CC EMBL; U46994; AAB19038.1; JOINED.
CC EMBL; U46995; AAB19038.1; JOINED.
CC EMBL; U46996; AAB19038.1; JOINED.
CC EMBL; U46997; AAB19038.1; JOINED.
CC EMBL; U46998; AAB19038.1; JOINED.
CC EMBL; U46999; AAB19038.1; JOINED.
CC EMBL; U47000; AAB19038.1; JOINED.
CC EMBL; U47001; AAB19038.1; JOINED.
CC EMBL; U47002; AAB19038.1; JOINED.
CC EMBL; U47003; AAB19038.1; JOINED.
CC EMBL; U47004; AAB19038.1; JOINED.
CC EMBL; U46960; AAB19039.1; JOINED.
CC EMBL; U46961; AAB19039.1; JOINED.
CC EMBL; U46962; AAB19039.1; JOINED.
CC EMBL; U46963; AAB19039.1; JOINED.
CC EMBL; U46964; AAB19039.1; JOINED.
CC EMBL; U46965; AAB19039.1; JOINED.
CC EMBL; U46966; AAB19039.1; JOINED.
CC EMBL; U46967; AAB19039.1; JOINED.
CC EMBL; U46968; AAB19039.1; JOINED.
CC EMBL; U46969; AAB19039.1; JOINED.
CC EMBL; U46970; AAB19039.1; JOINED.
CC EMBL; U46971; AAB19039.1; JOINED.
CC EMBL; U46972; AAB19039.1; JOINED.
CC EMBL; U46973; AAB19039.1; JOINED.
CC EMBL; U46974; AAB19039.1; JOINED.
CC EMBL; U46975; AAB19039.1; JOINED.
CC EMBL; U46976; AAB19039.1; JOINED.
CC EMBL; U46977; AAB19039.1; JOINED.
CC EMBL; U46978; AAB19039.1; JOINED.
CC EMBL; U46979; AAB19039.1; JOINED.
CC EMBL; U46980; AAB19039.1; JOINED.
CC EMBL; U46981; AAB19039.1; JOINED.
CC EMBL; U46982; AAB19039.1; JOINED.
CC EMBL; U46983; AAB19039.1; JOINED.

```

DR EMBL; U46984; AAB19039.1; JOINED.
DR EMBL; U46985; AAB19039.1; JOINED.
DR EMBL; U46986; AAB19039.1; JOINED.
DR EMBL; U46987; AAB19039.1; JOINED.
DR EMBL; U46988; AAB19039.1; JOINED.
DR EMBL; U46989; AAB19039.1; JOINED.
DR EMBL; U46990; AAB19039.1; JOINED.
DR EMBL; U46991; AAB19039.1; JOINED.
DR EMBL; U46992; AAB19039.1; JOINED.
DR EMBL; U46993; AAB19039.1; JOINED.
DR EMBL; U46994; AAB19039.1; JOINED.
DR EMBL; U46995; AAB19039.1; JOINED.
DR EMBL; U46996; AAB19039.1; JOINED.
DR EMBL; U46997; AAB19039.1; JOINED.
DR EMBL; U46998; AAB19039.1; JOINED.
DR EMBL; U47000; AAB19039.1; JOINED.
DR EMBL; U47001; AAB19039.1; JOINED.
DR EMBL; U47002; AAB19039.1; JOINED.
DR EMBL; U47003; AAB19039.1; JOINED.
DR EMBL; AL034369; CAA22265.1; -
DR EMBL; AL109943; CAB89263.1; -
DR EMBL; AL136080; CAB96748.1; -
DR EMBL; AL031177; CAA20120.1; -
DR EMBL; L22763; AAA16338.1; -
DR PIR; A54122; CGHU6B.
DR Genew; HGNC:2208; COL4A6.
DR NIM; 303631; -
DR GO; GO:0005587; C:collagen type IV; NAS.
DR GO; GO:0005201; Extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; Extracellular matrix organization and bioge. .; NAS.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
DR KX Extracellular matrix; Connective tissue; Basement membranes;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.

Query Match 78.2%; Score 115; DB 1; Length 1691;
Best Local Similarity 68.0%; Pred. No. 2.le-08;
Matches 17; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TMPLFCNVNDVCFASRNDYSWL 25
Db 1521 TMPLFYCNINEVCHYARRNDKSYWL 1545

RESULT 11
CA24_MOUSE STANDARD; PRT; 1707 AA.
ID CA24_MOUSE
AC P08122; Q61375;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; PubMed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumaran G.,
RA Pihlajaniemi T., Kurkinen M.;

```

```

RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen.";
RL J. Biol. Chem. 264:6318-6324(1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
RA Deutmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RL chain and its comparison with the alpha 1(IV) chain.";
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
RT "cDNA and protein sequence of the NCI domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1(IV).";
RL FEBS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen.";
RL FEBS Lett. 206:229-32(1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RL alpha 2(IV) collagen gene.";
RL Nature 317:177-179(1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M23334; AAA51626.1; -
 DR EMBL; M23333; AAA51626.1; JOINED.
 DR EMBL; J04695; AAA50293.1; -
 DR EMBL; J04448; AAA37438.1; -
 DR EMBL; J04647; CAA28308.1; -
 DR EMBL; M15833; AAA37341.1; -
 DR EMBL; X04410; CAA27998.1; -
 DR EMBL; X02896; CAA26655.1; -
 DR EMBL; X02897; CAA51644.1; -
 DR EMBL; X02898; CAA26657.1; -
 DR EMBL; X02899; CAA26658.1; -
 DR EMBL; X02899; CAA26658.1; -
 DR PIR; A33526; A33526.
 DR GDI; MGI:89455; Col4a2.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 21.
 DR ProDom; PD000007; C1g_helix; 7.
 DR ProDom; PD003923; ProcollagenC4; 1.
 DR SMART; SM00111; C4; 2.
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Signal.
 FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.
 FT CHAIN 184 1707 TRIPLE-HELICAL REGION.
 FT DOMAIN 184 1479 NONHELICAL REGION (NCL).
 FT DOMAIN 1480 1707 OR 1588 (BY SIMILARITY).
 FT DISULFID 1499 1588 OR 1588 (BY SIMILARITY).
 FT DISULFID 1532 1585 BY SIMILARITY.
 FT DISULFID 1544 1550 OR 1700 (BY SIMILARITY).
 FT DISULFID 1607 1703 OR 1703 (BY SIMILARITY).
 FT DISULFID 1641 1700 BY SIMILARITY.
 FT DISULFID 1653 1660 BY SIMILARITY.
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1270 1270 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 1051 1051 P -> R (IN REF. 6).
 FT CONFLICT 1097 1097 S -> G (IN REF. 7).
 FT CONFLICT 1171 1171 G -> S (IN REF. 6).
 FT CONFLICT 1179 1179 P -> R (IN REF. 6).
 FT CONFLICT 1241 1241 Q -> E (IN REF. 6).
 FT CONFLICT 1328 1328 P -> A (IN REF. 6).
 FT CONFLICT 1573 1573 V -> L (IN REF. 4).
 FT CONFLICT 1623 1623 Y -> H (IN REF. 4).
 SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605PD508 CRC64;

Query Match 76.2%; Score 112; DB 1; Length 1707;
 Best Local Similarity 76.0%; Pred. No. 5.7e-08;
 Matches 19; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 TMPEFLCNVDNVCNFRNDYSYL 25
 DB 1538 TMPEFLCNPGDVCYASRNDKSYL 1562

RESULT 12
 CA24 HUMAN STANDARD; PRT; 1712 AA.
 AC P08572;

DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 GN COL4A2.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89056769; PubMed=3198637;
 RA Hostikka S.L., Tryggvason K.;
 RT "The complete primary structure of the alpha 2 chain of human type IV
 RL collagen and comparison with the alpha 1(IV) chain.";
 RL J. Biol. Chem. 263:19488-19493(1988).
 RN [2]
 RP SEQUENCE OF 1-1042 FROM N.A.
 RX MEDLINE=88151998; PubMed=3345760;
 RA TISSUE=Placenta;
 RX MEDLINE=87219158; PubMed=3582677;
 RA Hostikka S.L., Kurkinen M., Tryggvason K.;
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2
 RL chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
 RL (IV) but not with the collagenous domain or 3'-untranslated region.";
 RL FEBS Lett. 216:281-286(1987).
 RN [4]
 RP SEQUENCE OF 1451-1485 FROM N.A.
 RX MEDLINE=87092438; PubMed=3025878;
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
 RL alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
 RN [5]
 RP SEQUENCE OF 1486-1712 FROM N.A.
 RX MEDLINE=87250571; PubMed=2439508;
 RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
 RT "Duplication of type IV collagen COOH-terminal repeats and species-
 RL specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
 RL J. Biol. Chem. 262:9231-9238(1987).
 RN [6]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89034231; PubMed=3182844;
 RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RL collagen are divergently encoded on opposite DNA strands and have an
 RL overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [7]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89030632; PubMed=2846280;
 RA Poeschl E., Pollner R., Kuehn K.;
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
 RL basement membrane collagen type IV are arranged head-to-head and
 RL separated by a bidirectional promoter of unique structure.";
 RL EMBO J. 7:2667-2695(1988).
 RN [8]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX TISSUE=Skin;
 RX MEDLINE=93305049; PubMed=8317999;
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
 RT "Identification of a novel sequence element in the common promoter
 RL region of human collagen type IV genes, involved in the regulation of
 RL divergent transcription.";
 RL Biochem. J. 292:687-695(1993).
 RN [9]
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
 RC TISSUE=Placenta;
 RX MEDLINE=89005112; PubMed=2844531;
 RA Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the

carboxyterminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen.";

RT Eur. J. Biochem. 176:617-624(1988).

CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM) forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X05562; CAA29076.1; -

CC EMBL; X05610; CAA29098.1; -

CC EMBL; J02760; AAS58422.1; -

CC EMBL; M36963; AAS30399.1; -

CC EMBL; X12784; CAA31275.1; -

CC EMBL; J04217; AAS3097.1; -

CC PIR; A32024; CGH72B.

CC Genew; HGNC:2203; COL4A2.

CC MIM; 120090; -

CC GO; GO:0005587; C:collagen type IV; TAS.

CC GO; GO:0003201; F:extracellular matrix structural constituent; TAS.

CC GO; GO:0030198; P:extracellular matrix organization and biogenesis; NAS.

CC InterPro; IPR008161; Clg_helix.

CC InterPro; IPR008160; Collagen.

CC InterPro; IPR001442; Procollagn4_C.

CC Pfam; PF01413; C4; 2.

CC Pfam; PF01391; Collagen; 24.

CC ProDom; PD000007; Clg_helix; 7.

CC ProDom; PD003923; ProcollagnC4; 1.

CC SMART; SM00111; C4; 2.

CC KW Glycoprotein; Basement membrane; Collagen; Signal.

CC FT SIGNAL 1 25

CC FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

CC FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.

CC FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.

CC FT DISULFID 1485 1712 NONHELICAL REGION (NC1).

CC FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).

CC FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).

CC FT DISULFID 1549 1555 BY SIMILARITY.

CC FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).

CC FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).

CC FT DISULFID 1658 1665 BY SIMILARITY.

CC FT CARBOHYD 138 138 N-LINKED (GLCNAC...).

CC FT CONFLICT 471 471 R -> P (IN REF. 2).

CC FT CONFLICT 683 683 A -> G (IN REF. 2).

CC FT CONFLICT 1575 1575 M -> I (IN REF. 5).

CC FT CONFLICT 1663 1663 G -> H (IN REF. 9).

CC FT CONFLICT 1701 1701 H -> G (IN REF. 9).

CC SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match 76.2%; Score 112; DB 1; Length 1712;

Best Local Similarity 76.0%; Pred. No. 5.7e-08;

Matches 19; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 TWPFLFCNVNDYCNFASNDYSYWL 25

DB 1543 TWPFLYCNPDVCYVYASRNDKSYWL 1567

RESULT 13

CA44_RABIT STANDARD; PRT; 623 AA.

ID CA44_RABIT

AC P55787;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Collagen alpha 4(IV) chain (Fragment).

GN COL4A4.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Corneal endothelium;

RX MEDLINE=93054733; PubMed=1429714;

RA Kamagata Y., Mattei M.-G., Ninomiya Y.;

RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha 4 chain of basement membrane collagen type IV and assignment of the gene to the distal long arm of human chromosome 2.";

RL J. Biol. Chem. 267:23753-23758(1992).

RL -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; L01477; -; NOT ANNOTATED_CDS.

CC PIR; A45137; A45137.

CC InterPro; IPR008160; Collagen.

CC InterPro; IPR001442; Procollagn4_C.

CC Pfam; PF01413; C4; 2.

CC Pfam; PF01391; Collagen; 5.

CC ProDom; PD003923; ProcollagnC4; 1.

CC SMART; SM00111; C4; 2.

CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Cell adhesion.

CC FT NON_TER 1 1

CC FT DOMAIN <1 392 TRIPLE-HELICAL REGION.

CC FT DOMAIN 393 623 NONHELICAL REGION (NC1).

CC FT DISULFID 413 502 OR 499 (BY SIMILARITY).

CC FT DISULFID 446 499 OR 502 (BY SIMILARITY).


```
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38
FT CHAIN 39 1690
FT DOMAIN 39 64
FT DOMAIN 65 1459
FT DOMAIN 1460 1690
FT SITE 94 96
FT SITE 145 147
FT SITE 189 191
FT SITE 310 312
FT SITE 724 726
FT SITE 785 787
FT SITE 989 991
FT SITE 1206 1207
FT SITE 1212 1214
FT DISULFID 1480 1569
FT DISULFID 1513 1566
FT DISULFID 1525 1531
FT DISULFID 1588 1686
FT DISULFID 1622 1683
FT DISULFID 1634 1641
FT CARBOHYD 142 144
FT CARBOHYD 669 669
FT VARIANT 441 446
FT VARIANT 545 545
FT VARIANT 570 570
FT VARIANT 897 897
FT VARIANT 931 931
FT VARIANT 1004 1004
FT VARIANT 1030 1030
FT VARIANT 1201 1201
FT VARIANT 1402 1402
FT VARIANT 1572 1572
FT CONFLICT 1659 1660
FT SEQUENCE 1690 AA; 164095 MW; E1E72F283A72BAE CRC64;
Query Match 70.1%; Score 103; DB 1; Length 1690;
Best Local Similarity 60.0%; Pred. No. 1.1e-06;
Matches 15; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 1 TMAPLFCNVNDVCNFAASNDYSVWL 25
DQ 1519 TLFPAYCNHQVCHYAQRNDRSVWL 1543
RESULT 15
CA44_BOVIN STANDARD; PRT; 453 AA.
AC Q29432;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
```

Collagen alpha 4 (IV) chain (Fragment).

COL4A4.

Bos taurus (Bovine).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

NCBI_TaxID=9913;

[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.

TISSUE=Lens;

MEDLINE=92112769; PubMed=1370461;

Mariyama M., Kalluri R., Hudson B.G., Readers S.T.; "The alpha 4 (IV) chain of basement membrane collagen. Isolation of cDNAs encoding bovine alpha 4 (IV) and comparison with other type IV collagens"; J. Biol. Chem. 267:1253-1258(1992).

[2]

SEQUENCE OF 217-246.

MEDLINE=90202779; PubMed=2318822;

Gunwar S., Saus J., Noelken M.E., Hudson B.G.; "Glomerular basement membrane. Identification of a fourth chain, alpha 4, of type IV collagen."; J. Biol. Chem. 265:5466-5469(1990).

[3]

SEQUENCE OF 217-233.

MEDLINE=87222419; PubMed=2438283;

Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J., Hudson B.G.; "Localization of the Goodpasture epitope to a novel chain of basement membrane collagen."; J. Biol. Chem. 262:7874-7877(1987).

RT membrane collagen.";

CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).

CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are colocalized and present only in basement membranes of kidney, eye, cochlea, lung and brain.

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M77480; AAA30458.2; ALT_SEQ.

CC PIR; S18804; S18804.

CC InterPro; IPR008160; Collagen.

CC InterPro; IPR001442; Procollagn4_C.

CC Pfam; PF01413; C4; 2.

CC Pfam; PF01391; Collagen; 4.

CC ProDom; PD003923; ProcollagnC4; 1.

CC SMART; SM00111; C4; 2.

CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Extracellular matrix; Basement membrane; Collagen; Cell adhesion.

KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.

FT NON_TER 1
 FT DOMAIN <1 222
 FT DISULFID 223 453
 FT DISULFID 243 332
 FT DISULFID 276 329
 FT DISULFID 288 294
 FT DISULFID 351 449
 FT DISULFID 385 446
 FT DISULFID 397 404
 FT CONFLICT 219 219
 SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A6SBC1 CRC64;

Query Match 69.4%; Score 102; DB 1; Length 453;
 Best Local Similarity 60.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 TNPFLFCNVNVCNPNRNDYSYL 25
 DB 282 TLFFAYCNIHQVCHYARRNDRSYL 306

Search completed: April 5, 2004, 06:59:38
 Job time : 5.1477 secs

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|---------|--------------|
| 1 | 147 | 100.0 | 212 | 6 | Q28512 | macaca mulia |
| 2 | 147 | 100.0 | 245 | 4 | Q2NYC4 | homo sapien |
| 3 | 146 | 99.3 | 203 | 6 | Q23032 | sus scrofa |
| 4 | 146 | 99.3 | 203 | 6 | Q24682 | cyctolagus |
| 5 | 146 | 99.3 | 212 | 6 | Q24687 | ovis aries |
| 6 | 141 | 95.9 | 161 | 11 | Q61430 | mus musculus |
| 7 | 141 | 95.9 | 210 | 6 | Q28273 | canis famli |
| 8 | 141 | 95.9 | 225 | 6 | Q28271 | canis famli |
| 9 | 141 | 95.9 | 226 | 11 | Q391208 | mus musculus |
| 10 | 141 | 95.3 | 229 | 4 | Q2NF88 | homo sapien |
| 11 | 141 | 95.3 | 229 | 4 | Q2NYC5 | homo sapien |
| 12 | 141 | 95.9 | 246 | 11 | Q61435 | mus musculus |
| 13 | 141 | 95.9 | 979 | 13 | Q919K3 | gallus gall |
| 14 | 141 | 95.9 | 1075 | 4 | Q86X41 | homo sapien |
| 15 | 141 | 95.9 | 1621 | 4 | Q94H49 | homo sapien |
| 16 | 141 | 95.9 | 1569 | 11 | Q9QZ50 | mus musculus |

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25
|||||
Db 41 TMFFLCNVNDVCFASRNDYSYL 65

RESULT 2

Q9NYC4 PRELIMINARY; PRT; 245 AA.
AC Q9NYC4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tmsstatin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,
RA Srikreen M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.;
RT "Distinct anti-tumor properties of a type IV collagen domain derived
from basement membrane.";
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL; AF258351; AAP72632.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4 C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 100.0%; Score 147; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.7e-14; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25
|||||
Db 74 TMFFLCNVNDVCFASRNDYSYL 98

RESULT 3

Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47284; AAA91882.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.

InterPro; IPR001442; Procollagn4 C.
InterPro; IPR000504; RNA_rec_mot.
Pfam; PF01413; C4; 2.
ProDom; PD003923; ProcollagnC4; 1.
SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 203 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 99.3%; Score 146; DB 6; Length 203;
Best Local Similarity 96.0%; Pred. No. 2e-14; Indels 0; Gaps 0;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25
|||||
Db 41 TMFFLCNVNDVCFASRNDYSYL 65

RESULT 4

Q28682 PRELIMINARY; PRT; 203 AA.
AC Q28682
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47283; AAA91893.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4 C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 203 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816B4D9E30 CRC64;

Query Match 99.3%; Score 146; DB 6; Length 203;
Best Local Similarity 96.0%; Pred. No. 2e-14; Indels 0; Gaps 0;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25
|||||
Db 41 TMFFLCNVNDVCFASRNDYSYL 65

RESULT 5

Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).

```
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Kathamna I.,
RA Mason P.J., Fusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47282; AAA91904.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4 C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; Procollagn4 C.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 212 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 99.3%; Score 146; DB 6; Length 212;
Best Local Similarity 96.0%; Pred. No. 2.1e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 41 TMPFLFCNVNDVNCNFASRNDYSYWL 65

RESULT 6
Q61430 Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbauer I.;
RT "Cloning of the NC1 domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (iv) and
RT alpha5 (iv) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X82205; CAA57689.1; -.
DR PIR: S49488; S49488.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4 C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 161 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;
```

```
Query Match 95.9%; Score 141; DB 11; Length 161;
Best Local Similarity 92.0%; Pred. No. 9.1e-14;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 8 TMPFLFCNVNDVNCNFASRNDYSYWL 32

RESULT 7
Q28273 Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A3 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL: U50935; AAC48585.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4 C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; Procollagn4 C.
DR SMART: SM00111; C4; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 210 210
SQ SEQUENCE 210 AA; 31119E4CA823633D CRC64;

Query Match 95.9%; Score 141; DB 6; Length 210;
Best Local Similarity 92.0%; Pred. No. 1.2e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 51 TMPFLFCNVNDVNCNFASRNDYSYWL 75

RESULT 8
Q28271 Q28271 PRELIMINARY; PRT; 225 AA.
AC Q28271;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
```



```
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMC0111; C4; 2.
FT NON_TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5CID5 CRC64;

Query Match 95.9%; Score 141; DB 4; Length 229;
Best Local Similarity 92.0%; Pred. No. 1.3e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVNCNFSARNDSYWL 25
Db 59 TMPFLFCNINNVNCFASRNDSYWL 83

RESULT 12
Q61435 ID Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
laminae: Sequence, distribution, association with laminins, and
developmental switches.";
RL J. Cell Biol. 127:879-891 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z35166; CAA84529.1; -.
DR PIR; I48302; I48302.
DR MGD; MGI:104689; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMC0111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 95.9%; Score 141; DB 11; Length 246;
Best Local Similarity 92.0%; Pred. No. 1.4e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVNCNFSARNDSYWL 25
Db 75 TMPFLFCNINNVNCFASRNDSYWL 99

RESULT 13
Q919K3 ID Q919K3 PRELIMINARY; PRT; 979 AA.
AC Q919K3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE Collagen IV al chain (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
basal lamina.";
RL Dev. Biol. 0:0-0 (2000).
DR EMBL; AF239838; AAF44681.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 2.
DR SMART; SMC0111; C4; 2.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 95.9%; Score 141; DB 13; Length 979;
Best Local Similarity 92.0%; Pred. No. 5.8e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVNCNFSARNDSYWL 25
Db 809 TMPFLFCNINNVNCFASRNDSYWL 833

RESULT 14
Q86X41 ID Q86X41 PRELIMINARY; PRT; 1075 AA.
AC Q86X41;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047305; AAH47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SMC0111; C4; 2.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match 95.9%; Score 141; DB 4; Length 1075;
Best Local Similarity 92.0%; Pred. No. 6.4e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVNCNFSARNDSYWL 25
```

DB 905 TMEFLCINNVCFASRNDYSYWL 929

RESULT 15

| RESOUR | Q9H4R9 | PRELIMINARY; | PRT; 1621 AA. |
|--------|---|--------------|------------------------------------|
| AC | Q9H4R9; | | |
| AD | 01-MAR-2001 (T-EMBLrel. 16, Created) | | |
| DT | 01-MAR-2001 (T-EMBLrel. 16, Last sequence update) | | |
| DT | 01-OCT-2003 (T-EMBLrel. 25, Last annotation update) | | |
| DT | BA472K17.2 (Collagen type IV alpha 1) (fragment). | | |
| DE | COL4A1. | | |
| GN | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| OX | NCBI_taxID=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Bates K.; | | |
| RL | Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AL390755; CAC1153.1; -. | | |
| DR | GO; GO:0005581; C:collagen; IEA. | | |
| DR | GO; GO:0005201; F:extracellular matrix structural constituent; IEA. | | |
| DR | InterPro; IPR008161; Clg helix. | | |
| DR | InterPro; IPR008160; Collagen. | | |
| DR | InterPro; IPR001442; Procollagn4_C. | | |
| DR | Pfam; PF01413; C4; 2. | | |
| DR | Pfam; PF01391; Collagen; 24. | | |
| DR | ProDom; PDC00007; Clg helix; 5. | | |
| DR | ProDom; PDC03923; ProcollagnC4; 1. | | |
| DR | SMART; SM00111; C4; 2. | | |
| KW | Collagen. | | |
| FT | NON_TER | 1 | |
| SQ | SEQUENCE | 1621 AA; | 155705 MW; 73R5F901CD05DBA2 CRC64; |

Search completed: April 5, 2004, 07:03:57
Job time : 15.0121 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 22.5182 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221b-37

Perfect score: 147

Sequence: 1 TWPFLLFCNVNDVCFASRNDSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 147 | 100.0 | 25 | ADA20236 | Ada20236 T7 peptid |
| 2 | 147 | 100.0 | 79 | AAU75600 | AAU75600 Human typ |
| 3 | 147 | 100.0 | 79 | ADA20264 | Ada20264 Human tum |
| 4 | 147 | 100.0 | 88 | AAU75608 | AAU75608 Human typ |
| 5 | 147 | 100.0 | 88 | AAU75607 | AAU75607 Human typ |
| 6 | 147 | 100.0 | 88 | ADA20271 | Ada20271 Human tum |
| 7 | 147 | 100.0 | 88 | ADA20272 | Ada20272 Human tum |
| 8 | 147 | 100.0 | 124 | AAU75594 | AAU75594 Human typ |
| 9 | 147 | 100.0 | 124 | ADA20258 | Ada20258 Human typ |
| 10 | 147 | 100.0 | 132 | AAU75597 | AAU75597 Human typ |
| 11 | 147 | 100.0 | 132 | ADA20261 | Ada20261 Human tum |
| 12 | 147 | 100.0 | 191 | AAU75596 | AAU75596 Human typ |
| 13 | 147 | 100.0 | 191 | ADA20260 | Ada20260 Human tum |
| 14 | 147 | 100.0 | 211 | AAU75598 | AAU75598 Human typ |
| 15 | 147 | 100.0 | 211 | ABG79208 | ABG79208 Human GP |
| 16 | 147 | 100.0 | 218 | AAU75164 | AAU75164 Partial s |
| 17 | 147 | 100.0 | 218 | AAU75164 | AAU75164 Partial s |
| 18 | 147 | 100.0 | 218 | AAU756784 | AAU756784 Human alp |
| 19 | 147 | 100.0 | 218 | AAU756784 | AAU756784 Human alp |
| 20 | 147 | 100.0 | 232 | ADCL17697 | ADCL17697 Human typ |
| 21 | 147 | 100.0 | 244 | ABG79218 | ABG79218 Human typ |
| 22 | 147 | 100.0 | 244 | ABG79219 | ABG79219 Human typ |
| 23 | 147 | 100.0 | 244 | ABG79217 | ABG79217 Human typ |
| 24 | 147 | 100.0 | 244 | AAU75595 | AAU75595 Human typ |
| 25 | 147 | 100.0 | 244 | ADA20225 | Ada20225 Human typ |

| | | | | | | |
|----|-----|-------|------|---|------------|----------------------|
| 26 | 147 | 100.0 | 245 | 3 | AAU7567942 | AAU7567942 Human typ |
| 27 | 147 | 100.0 | 245 | 5 | AAU75589 | AAU75589 Human typ |
| 28 | 147 | 100.0 | 254 | 5 | AAU75598 | AAU75598 Human typ |
| 29 | 147 | 100.0 | 268 | 2 | AAU751993 | AAU751993 Type IV c |
| 30 | 147 | 100.0 | 268 | 3 | AAU75555 | AAU75555 Human alp |
| 31 | 147 | 100.0 | 1670 | 7 | ADA47063 | Ada47063 Human pro |
| 32 | 146 | 99.3 | 471 | 2 | AAU75163 | AAU75163 Partial s |
| 33 | 146 | 99.3 | 471 | 2 | AAU75163 | AAU75163 Partial s |
| 34 | 146 | 99.3 | 471 | 3 | AAU756783 | AAU756783 Bovine al |
| 35 | 146 | 99.3 | 471 | 3 | AAU756783 | AAU756783 Bovine al |
| 36 | 141 | 95.9 | 229 | 1 | AAU75524 | AAU75524 Complete |
| 37 | 141 | 95.9 | 229 | 3 | AAU7567943 | AAU7567943 Human typ |
| 38 | 141 | 95.9 | 229 | 5 | AAU75587 | AAU75587 Human typ |
| 39 | 141 | 95.9 | 229 | 6 | ADA20217 | Ada20217 Human typ |
| 40 | 141 | 95.9 | 229 | 7 | ADCL17695 | ADCL17695 Human typ |
| 41 | 141 | 95.9 | 260 | 2 | AAU751991 | AAU751991 Type IV c |
| 42 | 141 | 95.9 | 260 | 3 | AAU75553 | AAU75553 Human alp |
| 43 | 141 | 95.9 | 406 | 3 | AAU755169 | AAU755169 Lung.canc |
| 44 | 141 | 95.9 | 1669 | 4 | AAU40863 | AAU40863 Human pol |
| 45 | 141 | 95.9 | 1669 | 5 | ABB90760 | ABB90760 Human Tum |

ALIGNMENTS

RESULT 1

ADA20236
ID ADA20236 standard; peptide; 25 AA.

XX AC ADA20236;

XX DT 20-NOV-2003 (first entry)

XX DE T7 peptide related to human type IV collagen alpha and angiogenesis.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX KW metastasis; basement membrane organisation; type IV collagen network;

XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX KW cytosolic; gene therapy; T7 peptide; tumstatin; human;

XX KW type IV collagen alpha 3 chain.

XX OS Homo sapiens.

XX FN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX DR N-ESDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor

XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 53; Page 45; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments

XX CC with anti-angiogenic properties. The invention also relates to the DNA

XX CC sequences which encode the novel proteins. A wide variety of diseases are

XX CC the result of undesirable angiogenesis. The formation of new capillaries

XX CC from pre-existing vessels is essential for tumour growth and metastasis.

XX CC Basement membrane organisation is dependent on the assembly of a type IV

XX CC collagenous network which may occur through the C-terminal globular non-

XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

XX CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the T7 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 CC
 SQ Sequence 25 AA;

Query Match 100.0%; Score 147; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.3e-14;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFNSRNDYSYWL 25
 |||||
 DB 1 TMPFLFCNVNDVCFNSRNDYSYWL 25

RESULT 2
 AAU75600
 ID AAU75600 standard; protein; 79 AA.
 XX
 AC AAU75600;

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tum-5.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

PN WO200151523-A2.

PD 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.

XX Example 40; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues
 CC 54-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75589)

SQ Sequence 79 AA;

Query Match 100.0%; Score 147; DB 5; Length 79;

Best Local Similarity 100.0%; Pred. No. 4.6e-14;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFNSRNDYSYWL 25

DB 21 TMPFLFCNVNDVCFNSRNDYSYWL 45

RESULT 3

ADA20264

ID ADA20264 standard; protein; 79 AA.

XX

AC ADA20264;

DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-5 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX Homo sapiens.

OS WO2003059257-A2.

PN 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX
 PS Claim 94; SEQ ID NO 26; 240pp; English.
 XX

CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.
 XX

SQ Sequence 79 AA;

Query Match 100.0%; Score 147; DB 6; Length 79;
 Best Local Similarity 100.0%; Pred. No. 4.6e-14;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFLFCNVNDVCFASRNDYSYWL 25
 |||||
 DB 20 TMAPFLFCNVNDVCFASRNDYSYWL 44

RESULT 4
 AAU75608
 ID AAU75608 standard; protein; 88 AA.

XX AAU75608;

DT 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 82 /note= "Wild type Cys substituted with Ala"

XX WC200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX

DR WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.

XX Claim 41; Page 153; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
 CC consists of residues 5-126 of Tumstatin

XX Sequence 88 AA;

Query Match 100.0%; Score 147; DB 5; Length 88;
 Best Local Similarity 100.0%; Pred. No. 5.2e-14;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFLFCNVNDVCFASRNDYSYWL 25
 |||||
 DB 30 TMAPFLFCNVNDVCFASRNDYSYWL 54

RESULT 5
 AAU75607

ID AAU75607 standard; protein; 88 AA.

XX AAU75607;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.
 XX Homo sapiens.

XX WO200151523-A2.
PN 19-JUL-2001.
PD 08-JAN-2001; 2001WO-US000565.
XX 07-JAN-2000; 2000US-00479118.
PR 04-APR-2000; 2000US-00543371.
PR 21-JUL-2000; 2000US-00625191.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA Kalluri R;
PI WPI; 2002-188037/24.
DR A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX Claim 32; Page 152; 205pp; English.
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha5, alpha6, alphaV, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists
CC of residues 45-132 of Tumstatin
XX
SQ Sequence 88 AA;
Query Match 100.0%; Score 147; DB 5; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.2e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMAPFLFCNVNDVCFASRNDYSYWL 25
DB 30 TMAPFLFCNVNDVCFASRNDYSYWL 54

ADA20271
ID ADA20271 standard; protein; 88 AA.
XX AC ADA20271;
XX DT 20-NOV-2003 (first entry)
XX DE Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human;
XX tumstatin 45-132.
XX OS Homo sapiens.
XX PN WO2003059257-A2.
XX PD 24-JUL-2003.
XX PF 20-DEC-2002; 2002WO-US040938.
XX PR 21-DEC-2001; 2001US-00032221.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2003-597256/55.
XX DR N-PSDB; ADA20224.
XX PT New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX PS Claim 94; SEQ ID NO 33; 240pp; English.
XX CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
CC ID33) does not appear in the specification but was created by the indexer
CC from information given in the specification.
XX
SQ Sequence 88 AA;
Query Match 100.0%; Score 147; DB 6; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.2e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMAPFLFCNVNDVCFASRNDYSYWL 25
DB 29 TMAPFLFCNVNDVCFASRNDYSYWL 53

RESULT 7
ADA20272

ADA20272 standard; protein; 88 AA.
 ADA20272;
 20-NOV-2003 (first entry)
 Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.
 anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network; C-terminal globular non-collagenous domain; NCI; type IV collagen; cell surface receptor; integrin; angiogenic activity; protein synthesis; cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 5-125-C-A; mutant; mutein.
 Synthetic.
 Homo sapiens.
 OS
 OS
 OS
 PH Key Location/Qualifiers
 FT Misc-difference 81
 FT /note= "Wild-type Cys substituted by Ala at position 125
 FT of full-length tumstatin"
 XX
 PN WO2003059257-A2.
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PR 21-DEC-2001; 2001US-00032221.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PI Kalluri R;
 PI WPI; 2003-587256/55.
 DR
 XX
 PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX
 PS Claim 94; SEQ ID NO 34; 240pp; English.
 CC
 CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
 CC the "tumstatin" protein of the invention which was derived from the amino
 CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This
 CC sequence (Seq ID33) does not appear in the specification but was created
 CC by the indexer from information given in the specification.
 XX
 SQ Sequence 88 AA;

Query Match 100.0%; Score 147; DB 6; Length 88;
 Best Local Similarity 100.0%; Pred. No. 5.2e-14;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TMFFLFCNVNDVCFASRNDYSYWL 25
 |||||
 DB 29 TMFFLFCNVNDVCFASRNDYSYWL 53

RESULT 8
 AAU75594
 ID AAU75594 standard; protein; 124 AA.
 XX
 AC AAU75594;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.
 XX
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphabeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW tumstatin; angiogenesis; tumour; mutein; mutant.
 XX
 OS Homo sapiens.
 XX
 PN WO200151523-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 08-JAN-2001; 2001WO-US000565.
 XX
 PR 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX
 PI Kalluri R;
 PI WPI; 2002-188037/24.
 DR
 XX
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.
 XX
 PS Example 33; Page; 205pp; English.
 XX
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphabeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or

CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
CC 1-132 of Tumstatin. Note: The present sequence is not shown in the
CC specification but is derived from the wild type human Tumstatin sequence
CC given in figure 18A (see AAU75589)

XX Sequence 132 AA;

Query Match 100.0%; Score 147; DB 5; Length 132;

Best Local Similarity 100.0%; Pred. No. 8e-14;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25

Db 74 TMPFLFCNVNDVCFASRNDYSYWL 98

RESULT 11

ID ADA20261

AC ADA20261 standard; protein; 132 AA.

XX ADA20261;

XX ADA20261;

DT 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NCl; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytosstatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

OS Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
PT Claim 94; SEQ ID NO 23; 240pp; English.
XX This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NCl) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NCl domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of
CC the invention which was derived from the amino acid sequence of the alpha
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does
CC not appear in the specification but was created by the indexer from
CC information given in the specification.

XX Sequence 132 AA;

Query Match 100.0%; Score 147; DB 6; Length 132;

Best Local Similarity 100.0%; Pred. No. 8e-14;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25

Db 73 TMPFLFCNVNDVCFASRNDYSYWL 97

RESULT 12

ID AAU75596

AC AAU75596 standard; protein; 131 AA.

XX AAU75596;

XX AAU75596;

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NCl domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCl domain used in detecting and
XX treating disorders involving angiogenesis.
PT Example 32; Page; 205pp; English.
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NCl
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit


```
PI Saus J;
XX WPI; 2000-572094/53.
DR N-PSDB; AAS0367.
XX
XX Novel Goodpasture antigen binding proteins useful for diagnosing and
XX treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX
XX Claim 36; Page 151-152; 158pp; English.
XX
XX The present sequence is that of human recombinant Goodpasture antigen
XX (GP) Deltav, i.e. an alternative form of human GP resulting from splicing
XX out of exon V. The recombinant protein, lacking the Met-1 residue, was
XX expressed in bacterial pellets using modified vector pET15b carrying
XX GPDeltav cDNA (see AAS0367). The invention relates to novel Goodpasture
XX antigen binding proteins (GPBPs, see AAY55900-11), which bind to and
XX phosphorylate the unique N-terminal region of human GP, and which are
XX highly expressed in several autoimmune conditions. Claimed methods for
XX treating an autoimmune disorder, cell apoptosis or a tumour involve
XX modifying the expression or activity of GPBP, especially using a GP-
XX derived peptide, such as GPDeltav
XX
XX Sequence 211 AA;
SQ
Query Match 100.0%; Score 147; DB 3; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMPFLFCNVNVCNPFASRNDYSYWL 25
DB 73 TMPFLFCNVNVCNPFASRNDYSYWL 97
RESULT 15
ABG79208
ID ABG79208 standard; protein; 211 AA.
XX
XX AC ABG79208;
XX
XX DT 15-NOV-2002 (first entry)
XX
XX DE Human GP protein isoform GPDeltav.
XX
XX KW Goodpasture antigen binding protein; Goodpasture syndrome;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX
XX OS Homo sapiens.
XX
XX PN WO200261430-A2.
XX
XX PD 08-AUG-2002.
XX
XX PF 31-JAN-2002; 2002WO-EP001010.
XX
XX PR 31-JAN-2001; 2001US-0265249P.
XX
XX PA (SAUS/) SAUS J.
XX
XX PI Saus J;
XX
XX DR WPI; 2002-619280/66.
XX DR N-PSDB; ABS64491.
XX
XX Identifying candidate compounds for treating autoimmune conditions, e.g.
XX Goodpasture syndrome or lupus, comprises identifying compounds that
XX reduce phosphorylation of, or formation of conformational isomers of,
XX target proteins.
XX
XX Example 3; Page 199-200; 217pp; English.
```

```
XX The invention relates to identifying candidate compounds to treat an
XX autoimmune condition by identifying compounds that reduce phosphorylation
XX of a first target protein (I) (which is selected from Goodpasture antigen
XX binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)
XX domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
XX Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
XX Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
XX conformational isomers of the second target protein (II) (selected from
XX an alpha3 type IV collagen NC1 domain polypeptide and myelin basic
XX protein, MBP). Also included are (1) an isolated type IV collagen alpha3
XX NC1 domain conformational isomer, which has an amino acid sequence
XX identical to the wild type alpha3 type IV collagen NC1 domain, is
XX stabilised by disulphide bonds, and has a molecular weight in a non-
XX reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX a reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX type IV collagen alpha3 NC1 domain. The human gene for GPBP is located on
XX chromosome 5q13. The method is useful for treating autoimmune conditions,
XX such as Goodpasture syndrome, multiple sclerosis, systemic and cutaneous
XX lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
XX sequence represents an alpha3 type IV collagen non-collagenous (NC1)
XX domain (also known as the GP antigen) or an MBP isoform
XX
XX Sequence 211 AA;
SQ
Query Match 100.0%; Score 147; DB 5; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMPFLFCNVNVCNPFASRNDYSYWL 25
DB 73 TMPFLFCNVNVCNPFASRNDYSYWL 97
Search completed: April 5, 2004, 06:58:31
Job time : 22:5182 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 15.678 Seconds

(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-37

Perfect score: 147

Sequence: 1 TMAPFLFCNVNDVCFNFSRNDYSYL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 147 | 100.0 | 25 | 14 | US-10-032-221B-37 |
| 2 | 147 | 100.0 | 79 | 14 | US-10-032-221B-26 |
| 3 | 147 | 100.0 | 88 | 14 | US-10-032-221B-33 |
| 4 | 147 | 100.0 | 88 | 14 | US-10-032-221B-34 |
| 5 | 147 | 100.0 | 124 | 14 | US-10-032-221B-20 |
| 6 | 147 | 100.0 | 132 | 14 | US-10-032-221B-23 |
| 7 | 147 | 100.0 | 191 | 14 | US-10-032-221B-22 |
| 8 | 147 | 100.0 | 211 | 14 | US-10-270-877-46 |
| 9 | 147 | 100.0 | 211 | 14 | US-10-270-837-46 |
| 10 | 147 | 100.0 | 232 | 14 | US-10-206-699-304 |
| 11 | 147 | 100.0 | 244 | 14 | US-10-032-221B-10 |
| 12 | 141 | 95.9 | 229 | 14 | US-10-206-699-302 |
| 13 | 141 | 95.9 | 229 | 14 | US-10-032-221B-2 |
| 14 | 141 | 95.9 | 406 | 9 | US-09-925-302-507 |
| 15 | 141 | 95.9 | 1669 | 15 | US-10-372-683-8 |

| | | | | | | |
|----|-----|------|------|----|---------------------|-------------------|
| 16 | 139 | 94.6 | 25 | 14 | US-10-032-221B-38 | Sequence 38, Appl |
| 17 | 139 | 94.6 | 229 | 14 | US-10-206-699-306 | Sequence 206, App |
| 18 | 139 | 94.6 | 309 | 9 | US-09-925-297-496 | Sequence 496, App |
| 19 | 129 | 87.8 | 46 | 9 | US-09-864-761-48095 | Sequence 48095, A |
| 20 | 125 | 85.0 | 27 | 14 | US-10-032-221B-39 | Sequence 39, Appl |
| 21 | 120 | 81.6 | 1759 | 15 | US-10-369-493-7032 | Sequence 7032, Ap |
| 22 | 117 | 79.6 | 1744 | 15 | US-10-369-493-5832 | Sequence 5832, Ap |
| 23 | 115 | 78.2 | 142 | 9 | US-09-864-761-38021 | Sequence 38021, A |
| 24 | 115 | 78.2 | 228 | 14 | US-10-206-699-307 | Sequence 307, App |
| 25 | 114 | 77.6 | 227 | 14 | US-10-206-699-266 | Sequence 266, App |
| 26 | 112 | 76.2 | 227 | 14 | US-10-206-699-303 | Sequence 303, App |
| 27 | 112 | 76.2 | 227 | 14 | US-10-032-221B-6 | Sequence 6, Appli |
| 28 | 112 | 76.2 | 430 | 9 | US-09-925-302-518 | Sequence 518, App |
| 29 | 112 | 76.2 | 459 | 15 | US-10-331-496A-27 | Sequence 27, Appl |
| 30 | 112 | 76.2 | 459 | 15 | US-10-372-683-30 | Sequence 30, Appl |
| 31 | 112 | 76.2 | 1712 | 10 | US-09-961-403-9 | Sequence 9, Appli |
| 32 | 108 | 73.5 | 22 | 14 | US-10-206-699-265 | Sequence 265, App |
| 33 | 106 | 72.1 | 22 | 14 | US-10-206-699-267 | Sequence 267, App |
| 34 | 105 | 71.4 | 27 | 14 | US-10-032-221B-40 | Sequence 40, Appl |
| 35 | 104 | 70.7 | 18 | 14 | US-10-206-699-260 | Sequence 260, App |
| 36 | 103 | 70.1 | 231 | 14 | US-10-206-699-305 | Sequence 305, App |
| 37 | 101 | 68.7 | 27 | 14 | US-10-032-221B-42 | Sequence 42, Appl |
| 38 | 98 | 66.7 | 18 | 14 | US-10-206-699-259 | Sequence 259, App |
| 39 | 96 | 65.3 | 18 | 14 | US-10-206-699-261 | Sequence 261, App |
| 40 | 93 | 63.3 | 18 | 14 | US-10-206-699-254 | Sequence 254, App |
| 41 | 93 | 63.3 | 19 | 14 | US-10-032-221B-41 | Sequence 41, Appl |
| 42 | 91 | 61.9 | 22 | 14 | US-10-206-699-270 | Sequence 270, App |
| 43 | 89 | 60.5 | 15 | 14 | US-10-206-699-212 | Sequence 212, App |
| 44 | 89 | 60.5 | 20 | 14 | US-10-206-699-289 | Sequence 289, App |
| 45 | 89 | 60.5 | 20 | 14 | US-10-032-221B-29 | Sequence 29, Appl |

ALIGNMENTS

RESULT 1

US-10-032-221B-37
; Sequence 37, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T7 (amino acids 73-97 of SEQ ID NO:10)
US-10-032-221B-37

Query Match 100.0%; Score 147; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.1e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

, , CURRENT APPLICATION NUMBER: US/10/032,220
, , CURRENT FILING DATE: 2008-12-21
, , PRIOR APPLICATION NUMBER: PCT/US01/005655
, , PRIOR FILING DATE: 2001-01-08
, , PRIOR APPLICATION NUMBER: US 09/625,191
, , PRIOR FILING DATE: 2000-07-21
, , PRIOR APPLICATION NUMBER: US 09/543,371
, , PRIOR FILING DATE: 2000-04-04
, , PRIOR APPLICATION NUMBER: US 09/479,118
, , PRIOR FILING DATE: 2000-01-07
, , PRIOR APPLICATION NUMBER: US 09/335,224
, , PRIOR FILING DATE: 1999-06-17
, , PRIOR APPLICATION NUMBER: US 60/126,175
, , PRIOR FILING DATE: 1993-03-25
, , PRIOR APPLICATION NUMBER: US 60/089,689
, , PRIOR FILING DATE: 1998-06-17
, , NUMBER OF SEQ ID NOS: 58

```

; CURRENT FILING DATE: 2001-12-21

Db 73 TMAPLFCVNDVCFASRNDYSYWL 97

RESULT 7

US-10-032-221B-22

Sequence 22, Application US/10032221B

Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

CURRENT APPLICATION NUMBER: US/10/032,221B

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: PCT/US01/00565

PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: US 09/625,191

PRIOR FILING DATE: 2000-07-21

PRIOR APPLICATION NUMBER: US 09/543,371

PRIOR FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: US 09/479,118

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/126,175

PRIOR FILING DATE: 1999-03-25

PRIOR APPLICATION NUMBER: US 60/089,689

PRIOR FILING DATE: 1998-06-17

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 20

LENGTH: 124

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)

US-10-032-221B-20

Query Match 100.0%; Score 147; DB 14; Length 124;

Best Local Similarity 100.0%; Pred. No. 1e-13; 0; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0;

QY 1 TMAPLFCVNDVCFASRNDYSYWL 25

Db 73 TMAPLFCVNDVCFASRNDYSYWL 97

RESULT 6

US-10-032-221B-23

Sequence 23, Application US/10032221B

Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

CURRENT APPLICATION NUMBER: US/10/032,221B

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: PCT/US01/00565

PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: US 09/625,191

PRIOR FILING DATE: 2000-07-21

PRIOR APPLICATION NUMBER: US 09/543,371

PRIOR FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: US 09/479,118

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/126,175

PRIOR FILING DATE: 1999-03-25

PRIOR APPLICATION NUMBER: US 60/089,689

PRIOR FILING DATE: 1998-06-17

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 23

LENGTH: 132

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)

US-10-032-221B-23

Query Match 100.0%; Score 147; DB 14; Length 132;

Best Local Similarity 100.0%; Pred. No. 1.1e-13; 0; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0;

QY 1 TMAPLFCVNDVCFASRNDYSYWL 25

Db 73 TMAPLFCVNDVCFASRNDYSYWL 97

RESULT 7

US-10-032-221B-22

Sequence 22, Application US/10032221B

Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

CURRENT APPLICATION NUMBER: US/10/032,221B

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: PCT/US01/00565

PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: US 09/625,191

PRIOR FILING DATE: 2000-07-21

PRIOR APPLICATION NUMBER: US 09/543,371

PRIOR FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: US 09/479,118

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/126,175

PRIOR FILING DATE: 1999-03-25

PRIOR APPLICATION NUMBER: US 60/089,689

PRIOR FILING DATE: 1998-06-17

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 22

LENGTH: 191

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)

US-10-032-221B-22

Query Match 100.0%; Score 147; DB 14; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.6e-13; 0; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0;

QY 1 TMAPLFCVNDVCFASRNDYSYWL 25

Db 20 TMAPLFCVNDVCFASRNDYSYWL 44

RESULT 8

US-10-270-877-46

Sequence 46, Application US/10270877

Publication No. US20030049791A1

GENERAL INFORMATION:

APPLICANT: Saus, Juan

TITLE OF INVENTION: Goodpasture Binding Protein

FILE REFERENCE: 98-723-AD1

CURRENT APPLICATION NUMBER: US/10/270,877

CURRENT FILING DATE: 2002-10-11

PRIOR APPLICATION NUMBER: 09/512,563

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/121,483

PRIOR FILING DATE: 1999-02-24

NUMBER OF SEQ ID NOS: 63

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 46

LENGTH: 211

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GPDV

US-10-270-877-46

Query Match 100.0%; Score 147; DB 14; Length 211;

Best Local Similarity 100.0%; Pred. No. 1.8e-13; 0; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0;

QY 1 TMAPFCNVNDVNCNFASTRNDYSYWL 25
Db 73 TMAPFCNVNDVNCNFASTRNDYSYWL 97

RESULT 9
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match 100.0%; Score 147; DB 14; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFASTRNDYSYWL 25
Db 73 TMAPFCNVNDVNCNFASTRNDYSYWL 97

RESULT 10
US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 304
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 3 chain
US-10-206-699-304

Query Match 100.0%; Score 147; DB 14; Length 232;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFASTRNDYSYWL 25
Db 61 TMAPFCNVNDVNCNFASTRNDYSYWL 85

RESULT 11
US-10-032-221B-10
; Sequence 10, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-10

Query Match 100.0%; Score 147; DB 14; Length 244;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFASTRNDYSYWL 25
Db 73 TMAPFCNVNDVNCNFASTRNDYSYWL 97

RESULT 12
US-10-206-699-302
; Sequence 302, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 302
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
US-10-206-699-302

```

FEATURE:
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-507

Query Match      95.9%; Score 141; DB 9; Length 406;
Best Local Similarity 92.0%; Pred. No. 2.5e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
    |||||:|:|||||||
DB 236 TMPFLFCNINNVCNFASTRNDYSYWL 260
    |||||:|:|||||||

RESULT 15
US-10-372-683-8
; Sequence 8, Application US/10372683
; Publication No. US20040009171A1
; GENERAL INFORMATION:
; TITLE OF INVENTION:
; APPLICANT: GERRITSEN, MARY E.
; APPLICANT: PEALE JR., FRANKLIN V.
; APPLICANT: WU, THOMAS D.
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
; FILE REFERENCE: P1928R1P1
; CURRENT APPLICATION NUMBER: US/10/372,683
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 10/271,690
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/344,534
; PRIOR FILING DATE: 2001-10-18
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 8
; LENGTH: 1669
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-372-683-8

Query Match      95.9%; Score 141; DB 15; Length 1669;
Best Local Similarity 92.0%; Pred. No. 1e-11;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
    |||||:|:|||||||
DB 1499 TMPFLFCNINNVCNFASTRNDYSYWL 1523
    |||||:|:|||||||

Search completed: April 5, 2004, 07:36:06
Job time : 15.678 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 5.87167 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221B-37
Perfect score: 147
Sequence: 1 TNPFLFCNVNDVCFASNDYSYWL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:**
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCtUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 147 | 100.0 | 211 | 4 | US-09-512-563C-46 |
| 2 | 147 | 100.0 | 218 | 2 | US-08-399-889-25 |
| 3 | 147 | 100.0 | 218 | 3 | US-09-187-364-25 |
| 4 | 147 | 100.0 | 218 | 3 | US-09-439-897-4 |
| 5 | 147 | 100.0 | 268 | 4 | US-09-589-927-6 |
| 6 | 147 | 100.0 | 268 | 4 | US-09-277-665-6 |
| 7 | 147 | 100.0 | 268 | 4 | US-09-589-987-6 |
| 8 | 146 | 99.3 | 471 | 2 | US-08-399-889-24 |
| 9 | 146 | 99.3 | 471 | 3 | US-09-187-364-24 |
| 10 | 146 | 99.3 | 471 | 3 | US-09-439-897-2 |
| 11 | 141 | 95.9 | 260 | 4 | US-09-589-927-2 |
| 12 | 141 | 95.9 | 260 | 4 | US-09-277-665-2 |
| 13 | 141 | 95.9 | 260 | 4 | US-09-589-987-2 |
| 14 | 139 | 94.6 | 264 | 4 | US-09-589-927-10 |
| 15 | 139 | 94.6 | 264 | 4 | US-09-277-665-10 |
| 16 | 139 | 94.6 | 264 | 4 | US-09-589-987-10 |
| 17 | 115 | 78.2 | 260 | 4 | US-09-589-927-12 |
| 18 | 115 | 78.2 | 260 | 4 | US-09-277-665-12 |
| 19 | 115 | 78.2 | 260 | 4 | US-09-589-987-12 |
| 20 | 112 | 76.2 | 258 | 4 | US-09-589-927-4 |
| 21 | 112 | 76.2 | 258 | 4 | US-09-277-665-4 |
| 22 | 112 | 76.2 | 258 | 4 | US-09-589-987-4 |
| 23 | 103 | 70.1 | 260 | 4 | US-09-589-927-8 |
| 24 | 103 | 70.1 | 260 | 4 | US-09-277-665-8 |
| 25 | 103 | 70.1 | 260 | 4 | US-09-589-987-8 |
| 26 | 90 | 61.2 | 1694 | 1 | US-08-494-168-2 |
| 27 | 50.5 | 34.4 | 663 | 4 | US-09-194-468A-30 |

Sequence 3, Appli
Sequence 4389, Ap
Sequence 5407, Ap
Sequence 2, Appli
Sequence 6, Appli
Sequence 8623, Ap
Sequence 26, Appli
Sequence 381, App
Sequence 1, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 12, Appli
Sequence 12, Appli
Patent No. 5179198
Patent No. 5521296
Sequence 7, Appli
Sequence 13, Appli

28 47 32.0 186 2 US-08-766-551-3
29 46 31.3 400 4 US-03-134-000C-4389
30 45 30.6 410 4 US-09-543-881A-5407
31 44.5 30.3 704 3 US-08-792-832A-2
32 44 29.9 49 1 US-07-865-166A-6
33 44 29.9 313 4 US-09-489-039A-8623
34 44 29.9 525 4 US-09-549-519-26
35 43.5 29.6 164 4 US-09-634-338-381
36 43 29.3 103 1 US-08-271-562-1
37 43 29.3 103 1 US-08-087-007-3
38 43 29.3 103 2 US-08-696-777-1
39 43 29.3 103 3 US-08-483-433-3
40 43 29.3 103 5 PCT-US92-05920-3
41 41 29.3 105 3 US-09-591-435-12
42 43 29.3 128 6 5179198-1
43 43 29.3 128 6 5521296-1
44 43 29.3 160 4 US-09-668-673B-7
45 43 29.3 343 2 US-08-933-750C-13

ALIGNMENTS

RESULT 1
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512.563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 100.0%; Score 147; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNPFLFCNVNDVCFASNDYSYWL 25
DB 73 TNPFLFCNVNDVCFASNDYSYWL 97

RESULT 2
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT

; ORGANISM: Human
US-08-399-889-25

Query Match 100.0%; Score 147; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.8e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 47 TMPFLFCNVNDVCFASRNDYSYWL 71

RESULT 3

US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980

; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match 100.0%; Score 147; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.8e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 47 TMPFLFCNVNDVCFASRNDYSYWL 71

RESULT 4

US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558

; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match 100.0%; Score 147; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.8e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 47 TMPFLFCNVNDVCFASRNDYSYWL 71

RESULT 5

US-09-589-927-6
; Sequence 6, Application US/09589927

; Patent No. 6432706
; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match 100.0%; Score 147; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 97 TMPFLFCNVNDVCFASRNDYSYWL 121

RESULT 6

US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729

; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match 100.0%; Score 147; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 97 TMPFLFCNVNDVCFASRNDYSYWL 121

RESULT 7

US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140

; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match 100.0%; Score 147; DB 4; Length 268;

```
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||:|||||:|||||:
Db 97 TMAPFCNVNDVCFASRNDYSYWL 121

RESULT 8
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match 99.3%; Score 146; DB 2; Length 471;
Best Local Similarity 96.0%; Pred. No. 3.1e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||:|||||:|||||:
Db 300 TMAPFCNVNDVCFASRNDYSYWL 324

RESULT 9
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match 99.3%; Score 146; DB 3; Length 471;
Best Local Similarity 96.0%; Pred. No. 3.1e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||:|||||:|||||:
Db 300 TMAPFCNVNDVCFASRNDYSYWL 324

RESULT 10
US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 99.3%; Score 146; DB 3; Length 471;
Best Local Similarity 96.0%; Pred. No. 3.1e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||:|||||:|||||:
Db 300 TMAPFCNVNDVCFASRNDYSYWL 324

RESULT 11
US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 95.9%; Score 141; DB 4; Length 260;
Best Local Similarity 92.0%; Pred. No. 8.7e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||:|||||:|||||:
Db 90 TMAPFCNVNDVCFASRNDYSYWL 114

RESULT 12
US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2

Query Match 95.9%; Score 141; DB 4; Length 260;
Best Local Similarity 92.0%; Pred. No. 8.7e-13;
```


; CURRENT APPLICATION NUMBER: US/09/277,663

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.14528 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221B-38

Perfect score: 148

Sequence: 1 TMPFMFCNINNVCFASRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:**

1: PIR1:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----------|---------------------|
| 1 | 148 | 100.0 | 253 | 2 I48304 | collagen alpha 5(I) |
| 2 | 148 | 100.0 | 754 | 2 A55267 | collagen alpha 5(I) |
| 3 | 148 | 100.0 | 1691 | 1 S22917 | collagen alpha 5(I) |
| 4 | 145 | 98.0 | 161 | 2 S49488 | collagen alpha 3(I) |
| 5 | 145 | 98.0 | 246 | 2 I48302 | collagen alpha 3(I) |
| 6 | 145 | 98.0 | 258 | 2 B61228 | collagen alpha 1(I) |
| 7 | 145 | 98.0 | 1659 | 1 CGHU4B | collagen alpha 1(I) |
| 8 | 145 | 98.0 | 1659 | 1 CGMS4B | collagen alpha 1(I) |
| 9 | 140 | 94.6 | 471 | 2 A39024 | collagen alpha 3(I) |
| 10 | 139 | 93.9 | 220 | 2 A49736 | collagen alpha 3(I) |
| 11 | 139 | 93.9 | 1670 | 1 CGHU3B | collagen alpha 3(I) |
| 12 | 133 | 89.9 | 1752 | 2 A45407 | collagen alpha 3(I) |
| 13 | 129 | 87.2 | 1777 | 2 A54121 | collagen alpha 3(I) |
| 14 | 127 | 85.8 | 1783 | 2 S16366 | collagen alpha-4 c |
| 15 | 124 | 83.8 | 1758 | 2 T29350 | hypothetical prote |
| 16 | 124 | 83.8 | 1759 | 2 T29351 | collagen alpha 2(I) |
| 17 | 121 | 81.8 | 1744 | 2 S40991 | collagen alpha 1(I) |
| 18 | 120 | 81.1 | 261 | 2 A34476 | collagen alpha 2(I) |
| 19 | 113 | 76.4 | 1691 | 1 CGHU6B | collagen alpha 6(I) |
| 20 | 104 | 70.3 | 312 | 2 I48303 | collagen alpha 4(I) |
| 21 | 104 | 70.3 | 623 | 2 A45137 | collagen alpha 4(I) |
| 22 | 104 | 70.3 | 775 | 2 A61228 | collagen alpha 2(I) |
| 23 | 104 | 70.3 | 1690 | 1 CGHU1B | collagen alpha 4(I) |
| 24 | 104 | 70.3 | 1707 | 2 A33526 | collagen alpha 2(I) |
| 25 | 104 | 70.3 | 1712 | 1 CGHU2B | collagen alpha 2(I) |
| 26 | 103 | 69.6 | 453 | 2 S18804 | collagen alpha 4(I) |
| 27 | 96 | 64.9 | 1761 | 2 T13990 | collagen type IV a |
| 28 | 95 | 64.2 | 1775 | 2 A31893 | collagen alpha 1(I) |
| 29 | 67 | 45.3 | 79 | 2 C43928 | probable collagen |

30 58.5 39.5 58 2 B43928 probable collagen
31 50.5 34.1 257 2 H75419 hypothetical prote
32 50 33.6 235 2 T23439 hypothetical prote
33 49.5 33.4 886 2 T39081 hypothetical prote
34 48.5 32.8 388 2 T29364 hypothetical prote
35 48 32.4 523 2 T28727 hypothetical prote
36 48 32.4 531 2 T28222 hypothetical prote
37 48 32.4 1060 2 T30823 bumetanide sensiti
38 47.5 32.1 261 2 D96772 probable RING zinc
39 47.5 32.1 1743 2 T18279 multidrug resistan
40 47 31.8 249 2 T04939 hypothetical prote
41 47 31.8 356 2 T10514 probable stem brom
42 47 31.8 365 2 S45915 hypothetical prote
43 46.5 31.4 4981 2 T18489 hypothetical prote
44 46 31.1 98 1 F2NTK photoeystem II pro
45 46 31.1 118 2 T37269 hypothetical prote

ALIGNMENTS

RESULT 1

I48304

collagen alpha 5(IV) chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999

C:Accession: I48304; S47280

R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A>Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48304

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-253 <RES>

A:Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202

C:Superfamily: collagen alpha 1(IV) chain

Query Match 100.0%; Score 148; DB 2; Length 253;

Best Local Similarity 100.0%; Pred. No. 2,3e-13;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNINNVCFASRNDYSYWL 25

DB 83 TMPFMFCNINNVCFASRNDYSYWL 107

RESULT 2

A55267

collagen alpha 5(IV) chain - dog (fragment)

C:Species: Canis lupus familiaris (dog)

C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999

C:Accession: A55267

R:Zheng, K.; Thorner, P.S.; Marrano, P.; Baumal, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994

A>Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-

en type IV.

A:Reference number: A55267; MUID:94224868; PMID:8171024

A:Accession: A55267

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>

A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548

C:Superfamily: collagen alpha 1(IV) chain

Query Match 100.0%; Score 148; DB 2; Length 754;

Best Local Similarity 100.0%; Pred. No. 6,3e-13;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNINNVCFASRNDYSYWL 25

DB 591 TMPFMFCNINNVCFASRNDYSYWL 615

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25

Db 1521 TMPFMFCNNVNCNFCASRNDYSYWL 1545

RESULT 4
S49488

collagen alpha 3(IV) chain - mouse

C:Species: Mus musculus (house mouse)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999

C:Accession: S49488

R:Oberbaumer, I.

A:Description: Cloning of the NCI domains to the minor collagen IV chains of mouse via F

cells.

A:Reference number: S49487

A:Accession: S49488

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-161 <OBS>

A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G559916

C:Superfamily: collagen alpha 1(IV) chain

Query Match 98.0%; Score 145; DB 2; Length 161;

Best Local Similarity 96.0%; Pred. No. 3.9e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25

Db 8 TMPFLFCNNVNCNFCASRNDYSYWL 32

RESULT 5

I48302

collagen alpha 3(IV) chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997

C:Accession: I48302; S47278

R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48302

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-246 <RES>

A:Cross-references: EMBL:X35186; NID:G535197; PID:G535198

C:Superfamily: collagen alpha 1(IV) chain

Query Match

Best Local Similarity 98.0%; Score 145; DB 2; Length 246;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25

Db 75 TMPFLFCNNVNCNFCASRNDYSYWL 99

RESULT 6

B61228

collagen alpha 1(IV) chain - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999

C:Accession: B61228

R:Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.

Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991

A:Title: Cloning of alpha(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothe

A:Reference number: A61228; MUID:92010685; PMID:1717398

A:Accession: B61228

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-258 <YAM>

C:Superfamily: collagen alpha 1(IV) chain

Query Match

Best Local Similarity 98.0%; Score 145; DB 2; Length 258;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25

Db 88 TMPFLFCNNVNCNFCASRNDYSYWL 112

RESULT 7

CGH048

collagen alpha 1(IV) chain precursor - human

A:Alternate names: procollagen alpha 1(IV) chain

C:Species: Homo sapiens (man)

C:Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 07-Dec-1999

C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A1

R:Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 264, 13565-13571, 1989

A:Title: Structural organization of the gene for the alpha-1 chain of human type IV co

A:Reference number: S16876; MUID:89340433; PMID:2701944

A:Accession: S16876

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1669 <SOI1>

A:Cross-references: EMBL:J04217; GB:J05039; NID:G180759; PIDN:AAAS3098.1; PID:G180803

R:Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988

A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen a

A:Reference number: A92690; MUID:89034231; PMID:3182844

A:Accession: A32117

A:Molecule type: DNA

A:Residues: 1-28 <SOI2>

A:Cross-references: EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PID:G553233

R:Poschl, E.; Pollner, R.; Kuehn, K.

EMBO J. 7, 2687-2695, 1988

A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane

A:Reference number: S02738; MUID:89030632; PMID:2846280

A:Accession: S02738

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-6, 'L', 8-28 <POE>

R:Brazel, D.; Oberbaumer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.

Eur. J. Biochem. 168, 529-536, 1987

A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement m

A:Reference number: S00048; MUID:88029471; PMID:3311751

A:Accession: S00048

A:Molecule type: mRNA

A:Residues: 1-318, 'A', 320-944 <BRA1>

A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067

A:Accession: S25826

A:Molecule type: protein

R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.

Eur. J. Biochem. 152, 213-219, 1985

A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (75

A:Reference number: A23115; MUID:86004708; PMID:4043082

A:Accession: A23115

A:Molecule type: protein

A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

A:Experimental source: placenta

R:Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.

FEBS Lett. 225, 188-194, 1987

A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane (1

A:Reference number: S00207; MUID:88083584; PMID:3691802

A:Accession: S00207

A:Molecule type: mRNA

A:Residues: 244-530 <SOI3>

A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549
 R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
 EMO J. 12, 4795-4802, 1993
 A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
 A;Reference number: S39614; MUID:94038963; PMID:8223488
 A;Accession: S39614
 A;Molecule type: protein
 A;Residues: 371-554 <EBL>
 R;Babel, W.; Glanville, R.W.
 Eur. J. Biochem. 143, 545-556, 1984
 A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid se
 A;Reference number: A02863; MUID:85003629; PMID:6434307
 A;Accession: A02863
 A;Molecule type: protein
 A;Residues: 534-718 'D', 720-836 'Y', 838-841 'P', 843-903 'Q', 905-913 'K', 915-997 'K', 999-
 A;Experimental source: placenta
 R;Glanville, R.W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
 A;Title: Peptin fragments of human placental basement-membrane collagens showing interru
 A;Reference number: S16908; MUID:82005835; PMID:6792033
 A;Accession: A58517
 A;Molecule type: protein
 A;Residues: 534-537 'G', 539 'G', 541-542 'X', 544-553 '1389-1405', 'XX', 1408-1409 'X', 1411-14
 R;MacWhirt, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
 Biochemistry 22, 4940-4948, 1983
 A;Title: Isolation and characterization of pepsin-solubilized human basement membrane (b
 A;Reference number: S16910; MUID:84053346; PMID:6416291
 A;Accession: S16910
 A;Molecule type: protein
 A;Residues: 534-537 'G', 539 'G', 541-542 'G', 544-549 '939-940', 'M', 942-944 'V', 946 'X', 948-
 A;Experimental source: placenta
 R;Pihlajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; P
 J. Biol. Chem. 260, 7681-7687, 1985
 A;Title: CDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
 A;Reference number: S01466; MUID:85207819; PMID:2581569
 A;Accession: S01466
 A;Molecule type: mRNA
 A;Residues: 1256-1669 <PIH>
 A;Cross-references: EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PID:G180424
 R;Brinker, J.M.; Gudas, L.J.; Lojdi, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;
 Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
 A;Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
 A;Reference number: S16879; MUID:85216555; PMID:2582422
 A;Accession: S16879
 A;Molecule type: mRNA
 A;Residues: 1259-1669 <BRI>
 R;Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A;Reference number: A02864; MUID:85127033; PMID:2578961
 A;Accession: S19091
 A;Molecule type: protein
 A;Residues: 1435-1461 'H', 1463-1482 'X', 1484-1491, 1501-1514 'X', 1516-1519; 1534-1553 'X',
 R;Siebold, B.; Deutzmann, R.; Kuehn, K.
 Eur. J. Biochem. 176, 617-624, 1988
 A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxy-term
 A;Reference number: S02550; MUID:89005112; PMID:2844531
 A;Contents: annotation; disulfide bonds
 C;Genetics:
 A;Gene: GDB:COL4A1
 A;Cross-references: GDB:119791; OMIM:120130
 A;Map position: 13q34-13q34
 A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/
 1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
 C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2
 C;Associations among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
 C;Associations in the interrupted helical domain (with disulfide and desmosine cr
 C;Function:
 A;Description: structural component of extracellular basement membrane
 C;Superfamily: collagen alpha 1(IV) chain
 C;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication
 F;1-26/Domain: signal sequence #status predicted <SIG>

P;27-1669/Product: collagen alpha 1(IV) chain #status predicted <WAT>
 F;23-162/Domain: amino-terminal nonhelical, 75 <YSD>
 F;163-1440/Domain: interrupted helical <COL>
 F;414-452/Region: integrin binding #status experimental
 F;597-599/Region: cell attachment (R-G-D) motif
 F;917-919/Region: cell attachment (R-G-D) motif
 F;968-970/Region: cell attachment (R-G-D) motif
 F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NCL>
 F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTL>
 F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>
 F;217/Modified site: blocked amino end (Ala) (in mature form) #status experimental
 F;31,36,39,41,125,434,467,470/Disulfide bonds: interchain #status predicted
 F;45,48,78,90,129,156,172,217,228,231,277,295,298,322,343,361,460,463,497,527,540,543,5
 1081,1084,1099,1117,1132,1150,1165,1182,1185,1188,1235,1265,1283,1304,1319,1328,13
 F;45,48,78,90,129,156,172,217,228,231,277,295,298,322,343,361,460,463,497,527,543,573,582,6
 99,1117,1132,1150,1165,1182,1185,1206,1235,1265,1283,1304,1319,1328,1340,1356,1371
 F;45,63,75,84,87,96,102,105,108,111,117,120,123,138,141,147,150,153,159,167,178,181,184
 F;419,422,425,439,445,448,451,479,485,491,494,503,512,518,524,530,546,549,552,555,561,56
 9,745,748,751,754,763/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;126/Binding site: carboxylate (Asn) (covalent) #status experimental
 F;129/Modified site: allysine (Lys) #status predicted
 F;172,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical
 F;272,645,839/Modified site: 4-hydroxyproline (Pro) #status atypical
 F;446-447/Cleavage site: Gly-Ile (Gelatinase B) #status predicted
 F;766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899,9
 231,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,1256
 431,1427/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F;1120,1268/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental
 F;1120,1424/Modified site: 3-hydroxyproline (Pro) #status absent
 F;1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F;1505-1511,1616-1622/Disulfide bonds: #status predicted
 F;1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
 Query Match 98.0%; Score 145; DB 1; Length 1659;
 Best Local Similarity 96.0%; Pred. No. 3.5e-12;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TMPFMFCINNVNCFASRNDYSYWL 25
 DB 1499 TMPFLFCINNVNCFASRNDYSYWL 1523
 RESULT 8
 CGMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000
 C;Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A32003; A31766; S1
 R;Muthukumar, G.; Blumberg, B.; Kurkinen, M.
 J. Biol. Chem. 264, 6310-6317, 1989
 A;Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Dif
 A;Reference number: A33525; MUID:89197932; PMID:2703490
 A;Accession: A33525
 A;Molecule type: mRNA
 A;Residues: 1-1669 <MUT>
 A;Cross-references: EMBL:J04694; NID:G556296; PIDN:AAAS0292.1; PID:G556297
 R;Wood, L.; Theriault, N.; Vogeli, G.
 FEBS Lett. 227, 5-8, 1988
 A;Title: cDNA clones completing the nucleotide and derived amino acid sequence of the a
 A;Reference number: S01454; MUID:88112221; PMID:3338568
 A;Accession: S01454
 A;Molecule type: mRNA
 A;Residues: 1-185; 'L', 187-318; 'S', 320-368; 'L', 370-402; 'F', 404-480; 'L', 482-492; 'H', 494-
 A;Cross-references: EMBL:X06777
 R;Killen, P.D.; Burselo, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988
 A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen cha
 A;Reference number: A28066; MUID:88243724; PMID:3379041
 A;Accession: A28066
 A;Molecule type: mRNA
 A;Residues: 1-129 <K11>

A:Cross-references: EMBL:J03758; NID:G192669; PIDN:AAA37439.1; PID:G192670
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A:Reference number: A02864; MUID:85127033; PMID:2578961
 A:Accession: A02864
 A:Molecule type: mRNA
 A:Residues: 1276-1669 <OBE>
 A:Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
 Gene 43, 301-304, 1986
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
 A:Reference number: A25636; MUID:86301886; PMID:3755692
 A:Accession: A25636
 A:Molecule type: mRNA
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
 A:Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287
 A:Note: the authors translated the codon CAG for residue 1374 as Arg
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj
 J. Biol. Chem. 262, 8496-8499, 1987
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
 A:Reference number: A94680; MUID:87250460; PMID:3597383
 A:Accession: A29301
 A:Molecule type: mRNA
 A:Residues: 1441-1669 <KUR>
 A:Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115
 R:Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y.
 J. Biol. Chem. 263, 12310-12314, 1988
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
 A:Reference number: S19079; MUID:88315019; PMID:2842328
 A:Accession: S19079
 A:Molecule type: DNA
 A:Residues: 1-28 <K12>
 A:Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503
 R:Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
 J. Biol. Chem. 263, 19274-19277, 1988
 A:Title: Head-to-head arrangement of murine type IV collagen genes.
 A:Reference number: A92702; MUID:89066738; PMID:3198626
 A:Accession: A32003
 A:Molecule type: DNA
 A:Residues: 1-28 <KAY>
 A:Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449
 R:Burbello, P.D.; Martin, G.R.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
 A:Reference number: A94220; MUID:89071759; PMID:3200851
 A:Accession: A31766
 A:Molecule type: DNA
 A:Residues: 1-28 <BUR>
 A:Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
 J. Biol. Chem. 261, 6654-6657, 1986
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
 A:Reference number: S19094; MUID:86196099; PMID:3009468
 A:Accession: S19094
 A:Molecule type: DNA
 A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>
 A:Cross-references: EMBL:M13027
 R:Schuppan, D.; Timpi, R.; Glanville, R.W.
 FEBS Lett. 115, 297-300, 1980
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
 A:Reference number: S16509; MUID:80246483; PMID:6772473
 A:Accession: S16509
 A:Molecule type: protein
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123
 R:Schuppan, D.; Glanville, R.W.; Timpi, R.
 Eur. J. Biochem. 123, 505-512, 1982
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin
 A:Reference number: A25991; MUID:82186723; PMID:6804236
 A:Accession: A25991
 A:Molecule type: protein
 A:Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 100
 61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119,

A:Accession: B25991
 A:Molecule type: protein
 A:Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-123
 3, 'SP', 1266, 'IR', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 13
 R:Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpi, R.
 Eur. J. Biochem. 139, 401-410, 1984
 A:Title: Subunit structure and assembly of the globular domain of basement-membrane co
 A:Reference number: S17801; MUID:84132058; PMID:6658021
 A:Accession: S17801
 A:Molecule type: protein
 A:Residues: 1435-1443 <WEB>
 C:Genetics:
 A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
 A:Note: the list of introns may be incomplete
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular m
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
 F:28-162/Domain: 78 <7SD>
 F:163-1440/Domain: collagenous, triple helix <COL>
 F:597-599/Region: cell attachment (R-G-D) motif
 F:781-783/Region: cell attachment (R-G-D) motif
 F:911-913/Region: cell attachment (R-G-D) motif
 F:968-970/Region: cell attachment (R-G-D) motif
 F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1441-1552/Region: duplication
 F:1553-1669/Region: duplication
 F:31.36.39.41.434.467.470/Disulfide bonds: interchain #status predicted
 F:126/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:972,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,1090
 98,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxyp
 F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 Query Match 98.0%; Score 145; DB 1; Length 1669;
 Best Local Similarity 96.0%; Pred. No. 3.5e-12;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
 DB 1499 TMPFLFCNNVNCVFASRNDYSYWL 1523
 RESULT 9
 A39024
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Dec-1992 #sequence revision 04-Dec-1992 #text change 13-Aug-1999
 A:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
 R:Morrison, K.B.; Germino, G.G.; Reeders, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding th
 A:Reference number: A39024; MUID:91093146; PMID:1985905
 A:Accession: A39024
 A:Molecule type: mRNA
 A:Residues: 1-471 <MOR>
 A:Cross-references: EMBL:M63139; NID:G162886; PIDN:AAA62708.1; PID:G162887
 R:Butkowsky, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A:Reference number: S18432; MUID:87222419; PMID:2438283
 A:Accession: S20672
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-244 <BUT>
 R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collage
 A:Reference number: S17802; MUID:88330844; PMID:3417661
 A:Accession: S17802
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.

Best Local Similarity 88.0%; Pred. No. 3.6e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCFAASNDYSYWL 25
|||||:|||||:|||||:|||||:|||||:
DB 82 TMPFLFCNVNVCFAASNDYSYWL 106
|||||:|||||:|||||:|||||:|||||:

RESULT 11
CGHU3B
collagen alpha 3(IV) chain precursor, long splice form - human
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
C:Accession: A54763; A43928; A44043; A45971; A39786
J. Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.
R. Biol. Chem. 269, 23013-23017, 1994
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpressed
A:Reference number: A54763; MUID:94364994; PMID:8083201
A:Accession: A54763
A:Molecule type: mRNA
A:Residues: 1-1670 <MAR>
A:Cross-references: GB:X80031; NID:G577563; PID:G577564
A:Experimental source: kidney
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the
A:Reference number: A43928; MUID:92147878; PMID:1737849
A:Accession: A43928
A:Molecule type: mRNA
A:Residues: 1331-1524, 'I', 1526-1670 <TUR>
A:Cross-references: GB:M81379
A:Experimental source: kidney
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
A:Reference number: A44043; MUID:93015926; PMID:1400291
A:Accession: A44043
A:Molecule type: DNA; mRNA
A:Residues: 1386-1670 <QUT>
A:Cross-references: GB:M92993; NID:G177895; PIDN:AAA21610.1; PID:G177896
A:Note: sequence extracted from NCBI backbone (NCBIP:115597)
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A:Reference number: A44738; MUID:94274734; PMID:8006044
A:Contents: annotation; erratum; correction to intronic sequence in A44043
R:Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A:Reference number: A45971; MUID:93280184; PMID:8505332
A:Accession: A45971
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1427-1444 <BER>
A:Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly iden
R:Morrison, K.E.; Mariyama, M.; Yang-feng, T.L.; Reiders, S.T.
Am. J. Hum. Genet. 49, 545-554, 1991
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain o
A:Reference number: A39786; MUID:91353570; PMID:1882840
A:Accession: A39786
A:Molecule type: mRNA
A:Residues: 1453-1593, 'A', 1595-1670 <WOR>
A:Cross-references: GB:I55790; NID:G234419; PIDN:ABA19637.1; PID:G234419
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitop
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-q37
A:Introns: 1385/1; 1418/1; 1489/1; 1547/2; 1585/3; 1643/2 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w

A;Introns: 1385/1; 1418/1; 1488/1; 1541/2; 1585/3; 1643/2 #status incomplete
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with

C;keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;29-161/Domain: amino-terminal nonhelical, 7S <7SD>
F;162-1523/Region: interrupted helical
F;1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

F:1,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F:426/Binding site: carboxylate (Asn) (covalent) #status predicted
F:1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 85.8%; Score 127; DB 2; Length 1763;
Best Local Similarity 80.0%; Pred. No. 1.2e-09;

Matches 20; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSRNDYSYWL 25
|||||:|||||:|||||
Db 1587 TMPFLFCDVNNVCNYSRNDKSYWL 1611

RESULT 15

T29350
hypothetical protein F01G12.5a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T29350
R/Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A/Description: The sequence of C. elegans cosmid F01G12.
A/Reference number: Z20611
A/Accession: T29350
A/Status: preliminary; translated from GH/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1758 <WUX>
A/Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a
A/Experimental source: strain Bristol N2; clone F01G12
C/Genetics:
A/Gene: CESP:F01G12.5a
A/Map position: X
A/Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3
C/Superfamily: collagen alpha 1(IV) chain

Query Match 83.8%; Score 124; DB 2; Length 1758;
Best Local Similarity 80.0%; Pred. No. 3.le-09;
Matches 20; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSRNDYSYWL 25
|||||:|||||:|||||
Db 1585 TMPFLFCDVNNVCNYSRNDKSYWL 1609

Search completed: April 5, 2004, 07:05:36
Job time : 6.14528 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.1477 Seconds

(without alignments)

413.557 Million cell updates/sec

Title: US-10-032-221b-38

Perfect score: 148

Sequence: 1 TWPFMFCNNVNCNPFASRNDYSWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|--------------|---------------------|
| 1 | 148 | 100.0 | 754 | 1 CA54 CANFA | Q28247 canis faml |
| 2 | 148 | 100.0 | 1685 | 1 CA54 HUMAN | P29400 homo sapien |
| 3 | 145 | 98.0 | 1669 | 1 CA14 HUMAN | P02462 homo sapien |
| 4 | 145 | 98.0 | 1669 | 1 CA14 MOUSE | P02463 mus musculus |
| 5 | 140 | 94.6 | 1471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 6 | 139 | 93.9 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 7 | 127 | 85.8 | 1763 | 1 CA24 ASCSU | P27393 ascaris suu |
| 8 | 121 | 81.8 | 1758 | 1 CA14 CAEEL | P17139 caenorhabdi |
| 9 | 120 | 81.1 | 1758 | 1 CA24 CAEEL | P17140 caenorhabdi |
| 10 | 113 | 76.4 | 1691 | 1 CA64 HUMAN | Q14031 homo sapien |
| 11 | 104 | 70.3 | 623 | 1 CA44 RABIT | P55787 oryctolagus |
| 12 | 104 | 70.3 | 1690 | 1 CA44 HUMAN | P53420 homo sapien |
| 13 | 104 | 70.3 | 1707 | 1 CA24 MOUSE | P08122 mus musculus |
| 14 | 104 | 70.3 | 1712 | 1 CA24 HUMAN | P08572 homo sapien |
| 15 | 103 | 69.6 | 453 | 1 CA14 DROME | Q29442 bos taurus |
| 16 | 95 | 64.2 | 1775 | 1 CA14 BOVIN | P08120 drosophila |
| 17 | 49 | 33.1 | 333 | 1 AMR1 HUMAN | Q944x0 homo sapien |
| 18 | 49 | 33.1 | 344 | 1 AMR1 MOUSE | Q944x0 mus musculus |
| 19 | 48 | 32.4 | 1060 | 1 NKCL MANSE | Q25479 manduca sex |
| 20 | 47.5 | 32.1 | 1743 | 1 TAGC DTCDI | Q23868 dictyosteli |
| 21 | 47 | 31.8 | 386 | 1 MUM2 YEAST | P38236 saccharomyc |
| 22 | 47 | 31.8 | 976 | 1 HMDH GIBFU | Q12577 gibberella |
| 23 | 45.5 | 30.7 | 704 | 1 OE66 NPVAC | Q00704 autographa |
| 24 | 45 | 30.4 | 397 | 1 YMP7 YEAST | Q04359 saccharomyc |
| 25 | 45 | 30.4 | 867 | 1 SUL1 COTCO | Q90xb6 coturnix co |
| 26 | 45 | 30.4 | 870 | 1 SUL1 MOUSE | Q8k007 mus musculus |
| 27 | 45 | 30.4 | 871 | 1 SUL1 RAT | Q8v160 rattus norv |
| 28 | 45 | 30.4 | 871 | 1 SUL1 HUMAN | Q8v166 homo sapien |
| 29 | 44.5 | 30.1 | 433 | 1 TC01 HUMAN | P20061 homo sapien |
| 30 | 44.5 | 30.1 | 743 | 1 NU5C CARCG | Q9t156 carpenteria |
| 31 | 44 | 29.7 | 296 | 1 SAPR STRPU | P11761 strongyloce |
| 32 | 44 | 29.7 | 308 | 1 META SALTI | Q8z1w1 salmonella |
| 33 | 44 | 29.7 | 308 | 1 META SALTY | P37413 salmonella |

34 44 29.7 359 1 LLCB_SUNY3
35 44 29.7 2273 1 ABCR_HUMAN
36 43.5 29.4 301 1 Y664 METJA
37 43.5 29.4 318 1 RLA0 MAIZE
38 43.5 29.4 663 1 NM02 CHICK
39 43.5 29.4 1850 1 VIT2 CHICK
40 43 29.1 256 1 PRN3 HUMAN
41 43 29.1 319 1 Y997 CAEEL
42 43 29.1 395 1 NH10 CAEEL
43 43 29.1 456 1 YC13 ASTIO
44 43 29.1 464 1 SYE2 COXBU
45 43 29.1 484 1 C24B_PIG

ALIGNMENTS

RESULT 1
CA54 CANFA
ID CA54 CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; TISSUE=Kidney;
RX MEDLINE=9242486; PubMed=5171024;
RA Zheng K., Thorne P.S., Marrano P., Bauman R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire',
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
CC canine X-linked hereditary nephritis (HN), a disease similar to
CC that in humans (also referred to as Alport syndrome) characterized
CC by progressive renal failure and neurosensory deafness.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC

Q08871 synechocyst
P78363 homo sapien
Q58078 methanococc
Q24573 zea mays (m
Q90611 gallus gall
P02845 gallus gall
P24158 homo sapien
Q09986 caenorhabdi
P41999 caenorhabdi
P14761 astasia lon
Q93b16 coxiella bu
P52649 s cytochrom

```
DR EMBL; U07888; AAB60258.1; -.
DR PIR; A55267; A55267.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD03923; Procollagen4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1
FT DOMAIN <1 530 TRIPLE-HELICAL REGION.
FT DOMAIN 531 >754 NON-HELICAL REGION (NC1).
FT DISULFID 552 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 ? OR 754 (BY SIMILARITY).
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON_TER 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA25B CRC64;

Query Match 100.0%; Score 148; DB 1; Length 754;
Best Local Similarity 100.0%; Pred. No. 5.7e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFA25BNDYSYWL 25
DB 591 TMPFMFCNNVNCNFA25BNDYSYWL 615

RESULT 2
CA54 HUMAN
ID CA54 HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316523; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.

RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D., Shear D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1572282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
```

RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
 RA van Oost B.A., Smeets H.J.M.; mutations in the COL4A5 gene of
 RT patients with Alport syndrome.";
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
 RN [14]
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;
 RP ARG-869; ARG-872 AND CYS-1241.
 RX MEDLINE=95322976; PubMed=7599631;
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
 RT "Detection of 12 novel mutations in the collagenous domain of the
 RT COL4A5 gene in Alport syndrome patients.";
 RL Hum. Mutat. 5:197-204(1995).
 RN [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE=96213750; PubMed=8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RA Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RT common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RP VARIANTS AS
 RX MEDLINE=96213754; PubMed=8651296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RT exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
 RA Giarras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RT syndrome.";
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RA Tverskaya S., Bobryna V., Tealykova F., Ignatova M.,
 RA Krasnopolskaya X., Evgrafov O.;
 RT "Substitution of A1498D in noncollagen domain of a5(IV) collagen
 RT chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RN [19]
 RP VARIANT AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RT syndrome and COL4A5 R1677Q.";
 RL Hum. Genet. 99:681-684(1997).
 RN [20]
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
 RP AND ASP-1596.
 RX MEDLINE=98112435; PubMed=9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 RT Alport syndrome.";
 RL Hum. Mutat. 1:S106-S109(1998).
 RN [21]
 RP VARIANTS AS VAL-420; 456-PRO-458 DEL; ASP-573; ASP-624; ASP-635;
 RP 802-GLY-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
 RP ARG-1156; GLU-1261; SER-1337 AND ARG-1649.
 RX MEDLINE=99063529; PubMed=9848783;
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,

RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected
 RT Alport syndrome using PCR and direct DNA sequencing.";
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
 RN [22]
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
 RX MEDLINE=20030197; PubMed=10561141;
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
 RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
 RT "Detection of mutations in the COL4A5 gene in over 90% of male
 RT patients with X-linked Alport's syndrome by RT-PCR and direct
 RT sequencing.";
 RL Am. J. Kidney Dis. 34:854-862(1999).
 RN [23]
 RP VARIANT AS ARG-822.
 RN
 Query Match 100.0%; Score 148; DB 1; Length 1685;
 Best Local Similarity 100.0%; Pred. No. 1.3e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWPFMFCNNNNVNFASRNDYSYWL 25
 DB 1515 TWPFMFCNNNNVNFASRNDYSYWL 1539
 RESULT 3
 CA14_HUMAN STANDARD; PRT; 1669 AA.
 ID CA14_HUMAN
 AC P02462;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=69340433; PubMed=2701944;
 RA Soiminen R., Huotari M., Garguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RT type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [2]
 RP SEQUENCE OF 46-1257 FROM N.A.
 RX TISSUE=Placenta;
 RX MEDLINE=88083584; PubMed=3691802;
 RA Soiminen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RT "Complete primary structure of the alpha 1-chain of human basement
 RT membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [3]
 RP SEQUENCE OF 1-943 FROM N.A.
 RX TISSUE=Placenta;
 RX MEDLINE=88029471; PubMed=3311751;
 RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,
 RA Deutzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 RT basement membrane collagen (type IV) reveals 21 non-triplet
 RT interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN [4]
 RP SEQUENCE OF 28-243.
 RX MEDLINE=86004708; PubMed=4043082;
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking
 RT region (7S domain) of the alpha 1 (IV) chain of human basement
 RT membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN [5]

SEQUENCE OF 534-1447.
RX MEDLINE=85003629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RT "Structure of human-basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1(IV) chain.";
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85207819; PubMed=2581969;
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA "Chung M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Solinen R., Huotari M., Hestikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE-Placenta; PubMed=2844531;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire',
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Lysines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26576; AAA53098.1; JOINED.
CC EMBL; J04217; AAA53098.1; JOINED.
CC EMBL; M26550; AAA53098.1; JOINED.

DR EMBL; M26540; AAA53098.1; JOINED.
DR EMBL; M26542; AAA53098.1; JOINED.
DR EMBL; M26543; AAA53098.1; JOINED.
DR EMBL; M26544; AAA53098.1; JOINED.
DR EMBL; M26545; AAA53098.1; JOINED.
DR EMBL; M26546; AAA53098.1; JOINED.
DR EMBL; M26547; AAA53098.1; JOINED.
DR EMBL; M26537; AAA53098.1; JOINED.
DR EMBL; M26538; AAA53098.1; JOINED.
DR EMBL; M26548; AAA53098.1; JOINED.
DR EMBL; M26549; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.
DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.
DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA68698.1; -.
DR EMBL; X05561; CAA29075.1; -.
DR EMBL; M10940; AAA52006.1; -.
DR EMBL; M11315; AAA52042.1; -.
DR PIR; S16876; CGH04B.
DR Genew; HGNC:2202; COL4A1.
DR MIM; 120130; -.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4_2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; Clg_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
DR SIGNAL 27
DR PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
DR CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
DR DOMAIN 173 1440 TRIPLE-HELICAL REGION.
DR DOMAIN 1441 1669 NONHELICAL REGION (NC1).
DR DOMAIN 126 126 N-LINKED (GLCNAC. .).
DR CARBOHYD 126 1551 OR 1548.
DR DISULFID 1460 1551 OR 1548.
DR DISULFID 1493 1548 OR 1551.
DR DISULFID 1505 1511 OR 1662.
DR DISULFID 1570 1665 OR 1662.
DR DISULFID 1604 1662 OR 1662.
DR DISULFID 1616 1622 OR 1622.
DR CONFLICT 237 238 SG -> KE (IN REF. 4).
DR CONFLICT 241 241 G -> K (IN REF. 4).
DR CONFLICT 319 319 Q -> A (IN REF. 3).
DR CONFLICT 719 719 N -> D (IN REF. 5).
DR CONFLICT 837 837 D -> Y (IN REF. 5).

```
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9B8A4 CRC64;

Query Match 98.0%; Score 145; DB 1; Length 1669;
Best Local Similarity 96.0%; Pred. No. 3.2e-12;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Cv 1 TMPFMFCNNVNCNPNFASNDYSYL 25
Db 1499 TMPFMFCNNVNCNPNFASNDYSYL 1523

RESULT 4
CA14 MOUSE
ID CA14 MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1996 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 35, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagens.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009469;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbello P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM) forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; J03758; AAA37439.1; -
CC ENBL; M23333; AAA51625.1; -
CC ENBL; J04694; AAA50292.1; -
CC ENBL; X06777; CAA29946.1; -
CC ENBL; X02201; CAA36132.1; -
CC ENBL; M15832; AAA37340.1; -
CC ENBL; M14042; AAA37342.1; -
CC ENBL; M12879; AAA37343.1; -
CC ENBL; M13024; -; NOT_ANNOTATED_CDS.
CC ENBL; M13025; -; NOT_ANNOTATED_CDS.
CC ENBL; M13027; AAA37344.1; -
CC ENBL; M13027; AAA37345.1; -
CC ENBL; M13027; AAA37346.1; -
CC ENBL; J04448; AAA37437.1; -
CC PIR; A35255; CGMS4B.
CC MGD; MGI:188454; Col4a1.
CC GO; GO:0005604; C:basement membrane; IDA.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagen4_C.
CC Pfam; PF01413; C4; 2.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X80031; CAA56335.1; --
 DR EMBL; AJ288487; CAC36101.1; JOINED.
 DR EMBL; AJ288488; CAC36101.1; JOINED.
 DR EMBL; AJ288489; CAC36101.1; JOINED.
 DR EMBL; AJ288490; CAC36101.1; JOINED.
 DR EMBL; AJ288491; CAC36101.1; JOINED.
 DR EMBL; AJ288492; CAC36101.1; JOINED.
 DR EMBL; AJ288493; CAC36101.1; JOINED.
 DR EMBL; AJ288494; CAC36101.1; JOINED.
 DR EMBL; AJ288495; CAC36101.1; JOINED.
 DR EMBL; AJ288496; CAC36101.1; JOINED.
 DR EMBL; AJ288497; CAC36101.1; JOINED.
 DR EMBL; AJ288498; CAC36101.1; JOINED.
 DR EMBL; AJ288499; CAC36101.1; JOINED.
 DR EMBL; AJ288500; CAC36101.1; JOINED.
 DR EMBL; AJ288501; CAC36101.1; JOINED.
 DR EMBL; AJ288502; CAC36101.1; JOINED.
 DR EMBL; AJ288503; CAC36101.1; JOINED.
 DR EMBL; AJ288504; CAC36101.1; JOINED.
 DR EMBL; AJ288505; CAC36101.1; JOINED.
 DR EMBL; AJ288506; CAC36101.1; JOINED.
 DR EMBL; AJ288507; CAC36101.1; JOINED.
 DR EMBL; AJ288508; CAC36101.1; JOINED.
 DR EMBL; AJ288509; CAC36101.1; JOINED.
 DR EMBL; AJ288510; CAC36101.1; JOINED.
 DR EMBL; AJ288511; CAC36101.1; JOINED.
 DR EMBL; AJ288512; CAC36101.1; JOINED.
 DR EMBL; AJ288513; CAC36101.1; JOINED.
 DR EMBL; AJ288514; CAC36101.1; JOINED.
 DR EMBL; AJ288515; CAC36101.1; JOINED.
 DR EMBL; AJ288516; CAC36101.1; JOINED.
 DR EMBL; AJ288517; CAC36101.1; JOINED.
 DR EMBL; AJ288518; CAC36101.1; JOINED.
 DR EMBL; AJ288519; CAC36101.1; JOINED.
 DR EMBL; AJ288520; CAC36101.1; JOINED.
 DR EMBL; AJ288521; CAC36101.1; JOINED.
 DR EMBL; AJ288522; CAC36101.1; JOINED.
 DR EMBL; AJ288523; CAC36101.1; JOINED.
 DR EMBL; AJ288524; CAC36101.1; JOINED.
 DR EMBL; AJ288525; CAC36101.1; JOINED.
 DR EMBL; AJ288526; CAC36101.1; JOINED.
 DR EMBL; AJ288527; CAC36101.1; JOINED.
 DR EMBL; AJ288528; CAC36101.1; JOINED.
 DR EMBL; AJ288529; CAC36101.1; JOINED.
 DR EMBL; AJ288530; CAC36101.1; JOINED.
 DR EMBL; AJ288531; CAC36101.1; JOINED.
 DR EMBL; AJ288532; CAC36101.1; JOINED.
 DR EMBL; AJ288533; CAC36101.1; JOINED.
 DR EMBL; AJ288534; CAC36101.1; JOINED.
 DR EMBL; AJ288535; CAC36101.1; JOINED.
 DR EMBL; AJ288536; CAC36101.1; JOINED.

Query Match 93.9%; Score 139; DB 1; Length 1670;
 Best Local Similarity 88.0%; Pred. No. 2.2e-11;
 Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFMFCNNVNCNFSASNDYSYL 25
 DB 1499 TWPFMFCNNVNCNFSASNDYSYL 1523

RESULT 7

CA24_ACSU
 ID CA24_ACSU STANDARD; PRT; 1763 AA.
 AC P27393;
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 OS Ascaris suum (pig roundworm) (Ascaris lumbricoidea).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS I AND II).
 RX MEDLINE=91340768; PubMed=1714907;
 RA Pettitt J., Kingston I.B.;
 RT "The complete primary structure of a nematode alpha 2(IV) collagen
 and the partial structural organization of its gene";
 RL J. Biol. Chem. 266:16149-16156(1991).
 CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
 CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
 CC Type IV collagen forms a mesh-like network linked through
 CC intermolecular interactions between 7S domains and between NC1
 CC domains.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=I;
 CC IsoId=P27393-1; Sequence=Displayed;
 CC Name=II;
 CC IsoId=P27393-2; Sequence=VSP_001159;
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the
 G-X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67507; AAA18014.1; --
 DR PIR; S16366; S16366.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 25.
 DR ProDom; PD000007; Clg helix; 6.
 DR ProDom; PD003923; ProcollagenC4; 1.
 DR SMART; SM00111; C4; 2.
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 DR Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.
 FT DOMAIN 27 42 7S DOMAIN.
 FT DOMAIN 43 1529 TRIPLE-HELICAL REGION.
 FT DOMAIN 1530 1763 NONHELICAL REGION (NC1).
 FT DISULFID 1548 1637 OR 1634 (BY SIMILARITY).
 FT DISULFID 1581 1634 OR 1637 (BY SIMILARITY).
 FT DISULFID 1593 1599 BY SIMILARITY.
 FT DISULFID 1656 1752 OR 1749 (BY SIMILARITY).
 FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).
 FT DISULFID 1702 1709 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 249 249 O-LINKED (XYL. .) (GLYCOSAMINOGLYCAN)
 FT (IN ISOFORM II) (POTENTIAL)
 FT VARSPLIC 230 266 GSGQPGPGPGPGVFTSGAKTIIGPEGAPGNKGEK ->
 FT GDIGPAGPGPGPGPGFTGSGSIVGRHSGDKGVK (in

```
FT isoform II).
FT /FTID=VSP 001159.
SQ SEQUENCE 1763 AA; 168526 MW; 304F528BC06AAE0D CRC64;

Query Match      85.8%; Score 127; DB 1; Length 1763;
Best Local Similarity 80.0%; Pred. No. 1e-09;
Matches 20; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFASRNDYSYL 25
Db 1587 TMPFLFCDVNNVNCYASRNDKSYWL 1611

RESULT 8
CA14_CABEL STANDARD; PRT; 1758 AA.
AC P17139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CLB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Boulton A., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Karshaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R., Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2; PubMed=2793871;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
```



```
DR EMBL; U46984; AAB19039.1; JOINED.
DR EMBL; U46985; AAB19039.1; JOINED.
DR EMBL; U46986; AAB19039.1; JOINED.
DR EMBL; U46987; AAB19039.1; JOINED.
DR EMBL; U46988; AAB19039.1; JOINED.
DR EMBL; U46989; AAB19039.1; JOINED.
DR EMBL; U46990; AAB19039.1; JOINED.
DR EMBL; U46991; AAB19039.1; JOINED.
DR EMBL; U46992; AAB19039.1; JOINED.
DR EMBL; U46993; AAB19039.1; JOINED.
DR EMBL; U46994; AAB19039.1; JOINED.
DR EMBL; U46995; AAB19039.1; JOINED.
DR EMBL; U46996; AAB19039.1; JOINED.
DR EMBL; U46997; AAB19039.1; JOINED.
DR EMBL; U46998; AAB19039.1; JOINED.
DR EMBL; U46999; AAB19039.1; JOINED.
DR EMBL; U47000; AAB19039.1; JOINED.
DR EMBL; U47001; AAB19039.1; JOINED.
DR EMBL; U47002; AAB19039.1; JOINED.
DR EMBL; U47003; AAB19039.1; JOINED.
DR EMBL; AL034369; CAA22265.1; -.
DR EMBL; AL105943; CAB96748.1; -.
DR EMBL; AL136080; CAB96748.1; -.
DR EMBL; AL031177; CAA20120.1; -.
DR EMBL; L22763; AAB16338.1; -.
DR PIR; A54122; CGHU6B.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -.
DR GO; GO:000587; C:collagen type IV; NAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46
Query Match 76.4%; Score 113; DB 1; Length 1691;
Best Local Similarity 72.0%; Pred No. 8.5e-08;
Matches 18; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 TMPEMFCNNVNCNPNASNDYSYL 25
| | | | | | | | | | | | | | | | | | | | |
Db 1521 TMPEFYCINEVCHVARRNDKSYL 1545
RESULT 11
CA44 RABBIT STANDARD; PRT; 623 AA.
AC P55787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4(IV) chain (fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
```

```
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
alpha 4 chain of basement membrane collagen type IV and assignment of
the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
CC !- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC !- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
CC !- SUBCELLULAR LOCATION: Cell surface (potential).
CC !- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the G-
X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC !- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC !- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these located in the NC1 domain, are conserved in all known type
IV collagens.
CC !- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; L01477; -; NOT ANNOTATED_CDS.
DR PIR; A45137; A45137.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; SM00111; C4; 2.
DR ProDom; SM003923; ProcollagnC4; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 1
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT DOMAIN 393 623 NONHELICAL REGION (NCL).
FT DISULFID 413 502 OR 499 (BY SIMILARITY).
FT DISULFID 446 499 OR 502 (BY SIMILARITY).
FT DISULFID 458 464 BY SIMILARITY.
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 OR 619 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SQ SEQUENCE 623 AA; 62393 MW; CCBC9BB31242FE82 CRC64;
Query Match 70.3%; Score 104; DB 1; Length 623;
Best Local Similarity 64.0%; Pred No. 5.5e-07;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 1 TMPEMFCNNVNCNPNASNDYSYL 25
| | | | | | | | | | | | | | | | | | | | |
Db 452 TLPEFYCNIHQVCHVARRNDKSYL 476
RESULT 12
CA44 HUMAN STANDARD; PRT; 1690 AA.
ID CA44 HUMAN
AC P53420;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 4(IV) chain precursor.
GN COL4A4.
OS Homo sapiens (Human).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eurelestomii;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=95014445; PubMed=7523402;
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Readers S.T.;
 RT "Complete primary structure of the human type IV collagen alpha 4 (IV)
 RT chain. Comparison with structure and expression of the other alpha
 RT (IV) chains.";
 RL J. Biol. Chem. 269:26172-26177(1994).
 RN [2]
 RP SEQUENCE OF 1-23 FROM N.A.
 RX MEDLINE=98196854; PubMed=9537506;
 RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
 RA Ninomiya Y.;
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
 RT 2q36.";
 RL FEBS Lett. 424:11-16(1998).
 RN [3]
 RP SEQUENCE OF 1219-1690 FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=93374047; PubMed=8365481;
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
 RT "cDNA isolation and partial gene structure of the human alpha 4 (IV)
 RT collagen chain.";
 RL FEBS Lett. 330:122-128(1993).
 RN [4]
 RP SEQUENCE OF 1407-1507 FROM N.A.
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kanagata Y., Mattei M.-G., Ninomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 RT alpha 4 chain of basement membrane collagen type IV and assignment of
 RT the gene to the distal long arm of human chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758(1992).
 RN [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 RN [6]
 RP VARIANT AS SER-1201.
 RX MEDLINE=95078927; PubMed=7987396;
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
 RA Pison Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
 RA Smeets H.J.M., Readers S.T.;
 RT "Identification of mutations in the alpha 3 (IV) and alpha 4 (IV)
 RT collagen genes in autosomal recessive Alport syndrome.";
 RL Nat. Genet. 8:77-82(1994).
 RN [7]
 RP VARIANT FBH GLU-897.
 RX MEDLINE=96379660; PubMed=8787673;
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
 RT "Benign familial hematuria due to mutation of the type IV collagen
 RT alpha4 gene.";
 RL J. Clin. Invest. 98:1114-1118(1996).
 RN [8]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99011253; PubMed=9792860;
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
 RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
 RT "Determination of the genomic structure of the COL4A4 gene and of
 RT novel mutations causing autosomal recessive Alport syndrome.";
 RL Am. J. Hum. Genet. 63:1329-1340(1998).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 alpha 6(IV), each of which can form a triple helix structure with
 2 other chains to generate type IV collagen network.
 -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
 colocalized and present only in basement membranes of kidney, eye,
 cochlea, lung and brain.
 -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the G-
 X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 -!- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 -!- DISEASE: Defects in COL4A4 are a cause of autosomal recessive
 Alport syndrome (AS) [MIM:203780], an hereditary disorder
 characterized by progressive glomerulonephritis, renal failure,
 hematuria, ocular abnormalities and deafness. The recessive form
 occurs equally between males and females.
 -!- DISEASE: Defects in COL4A4 are a cause of familial benign
 hematuria (FBH) [MIM:141200] or thin basement membrane disease.
 FBH is characterized by persistent hematuria, an electron
 microscopically detectable thin glomerular basement membrane (GBM)
 and an autosomal dominant mode of inheritance. Renal function
 remains normal. In children, differentiation between FBH and AS
 can be difficult, because both disorders are manifested by
 persistent hematuria and thin GBM at that age.
 -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 EMBL; X81053; CAA56943.1; -;
 EMBL; AB008496; BAA25065.1; -;
 EMBL; D17391; BAA04214.1; -;
 D R PIR; A55360; CGHUIB.
 D R Genew; HGNC:2206; COL4A4.
 D R MIM; 120131; -;
 D R MIM; 141200; -;
 D R MIM; 203780; -;
 D R InterPro; IPR008161; Clg helix.
 D R InterPro; IPR008160; Collagen.
 D R InterPro; IPR001442; Procollag4_C.
 D R Pfam; PF01413; C4; 2.
 D R Pfam; PF01391; Collagen; 21.
 D R ProDom; PD000007; Clg helix; 3.
 D R ProDom; PD003923; Procollag4; 1.
 D R SMART; SM00111; C4; 2.
 D R ProDom; PD003923; Procollag4; 1.
 D R Extracellular matrix; Connective tissue; Basement membrane; Repeat;
 D R Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
 D R Polymorphism; Alport syndrome.
 FT SIGNAL 1 38
 FT CHAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.
 FT DOMAIN 39 64 7S DOMAIN.
 FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
 FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1206 1207 CLEAVAGE (BY COLLAGENASE)
 FT (BY SIMILARITY).

```

RT RT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
RT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
RT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
RT DISULFID 1525 1531 BY SIMILARITY.
RT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
RT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
RT DISULFID 1634 1641 BY SIMILARITY.
RT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
RT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (POTENTIAL).
RT VARIANT 441 446 Missing (in AS).
RT VARIANT 545 545 /FTid=VAR_008148.
RT VARIANT 570 570 /FTid=VAR_008149.
RT VARIANT 570 570 E -> Q.
RT VARIANT 897 897 /FTid=VAR_008150.
RT VARIANT 931 931 G -> E (in FBH).
RT VARIANT 931 931 /FTid=VAR_001912.
RT VARIANT 1004 1004 A -> T.
RT VARIANT 1004 1004 /FTid=VAR_008151.
RT VARIANT 1030 1030 L -> P (in dSNP:1800517).
RT VARIANT 1030 1030 /FTid=VAR_008152.
RT VARIANT 1201 1201 G -> V (in AS).
RT VARIANT 1201 1201 /FTid=VAR_008153.
RT VARIANT 1402 1402 G -> S (in AS).
RT VARIANT 1402 1402 /FTid=VAR_001913.
RT VARIANT 1572 1572 P -> S.
RT VARIANT 1572 1572 /FTid=VAR_008154.
RT VARIANT 1572 1572 P -> L (in AS).
RT VARIANT 1572 1572 /FTid=VAR_008155.
RT CONFLICT 1559 1560 LQ -> FE (IN REF. 3).
SQ SEQUENCE 1690 AA; 164095 MW; E1E7F2F283A72BAAE CRC64;

Query Match 70.3%; Score 104; DB 1; Length 1690;
Best Local Similarity 64.0%; Pred. No. 1.5e-06;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TMPFWCINNVNCFASNDYSYWL 25
DB 1519 TLFPAYCNHQVCHYAQRNDRSYWL 1543

RESULT 13
CA24 MOUSE STANDARD; PRT; 1707 AA.
ID CA24 MOUSE
AC P08122; Q61375;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; PubMed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
RA Pihlajaniemi T., Kurkinen M.;
RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen."
RL J. Biol. Chem. 264:6318-6324 (1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes."
RL J. Biol. Chem. 263:19274-19277 (1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbaumer I., Glangville R.W.,
RA Deutzmann R., Timpl R., Kuehn K.;

```

```

"Structure of mouse type IV collagen. Amino-acid sequence of the C-
terminal 511-residue-long triple-helical segment of the alpha 2(IV)
chain and its comparison with the alpha 1(IV) chain."
Eur. J. Biochem. 157:49-56 (1986).
(4)
SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbaumer I., Kuehn K.;
RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of
collagen IV and its comparison with alpha 1(IV)."
FEBS Lett. 208:203-207 (1986).
(5)
SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
alpha 1(IV) and alpha 2(IV) collagen."
J. Biol. Chem. 262:8496-8499 (1987).
(6)
SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
the basement membrane (type IV) collagen."
FEBS Lett. 206:29-32 (1986).
(7)
SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
alpha 2(IV) collagen gene."
Nature 317:177-179 (1985).
(8)
SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
bidirectional promoter and a shared enhancer."
Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).
-!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the G-
X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical VS domain.
-!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M23334; AAA51626.1; .
DR EMBL; M23333; AAA51626.1; JOINED.
DR EMBL; J04695; AAA50293.1; .
DR EMBL; J04448; AAA37438.1; .
DR EMBL; X04647; CAA28308.1; .
DR EMBL; M15833; AAA37341.1; .
DR

```

```
DR EMBL; X04410; CAA27998.1; -
DR EMBL; X02896; CAA26655.1; -
DR EMBL; X02897; CAA51614.1; -
DR EMBL; X02898; CAA26657.1; -
DR EMBL; X02899; CAA26658.1; -
DR PIR; A3326; A3326.
DR MGI; MGI:88455; Col4a2.
DR CO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; C1g_helix; 7.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT PROPEP 26 183 COLLAGEN ALPHA 2 (IV) CHAIN.
FT CHAIN 184 1707 TRIPLE-HELICAL REGION.
FT DOMAIN 184 1479 NONHELICAL REGION (NC1).
FT DISULFID 1480 1707 OR 1585 (BY SIMILARITY).
FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).
FT DISULFID 1532 1585 BY SIMILARITY.
FT DISULFID 1544 1550 OR 1700 (BY SIMILARITY).
FT DISULFID 1607 1703 OR 1703 (BY SIMILARITY).
FT DISULFID 1641 1700 BY SIMILARITY.
FT DISULFID 1653 1660 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1270 1270 P -> R (IN REF. 6).
FT CONFLICT 1051 1051 G -> S (IN REF. 7).
FT CONFLICT 1097 1097 G -> S (IN REF. 6).
FT CONFLICT 1171 1171 P -> R (IN REF. 6).
FT CONFLICT 1179 1179 Q -> E (IN REF. 6).
FT CONFLICT 1241 1241 P -> A (IN REF. 4).
FT CONFLICT 1328 1328 V -> L (IN REF. 4).
FT CONFLICT 1573 1573 Y -> H (IN REF. 4).
FT CONFLICT 1623 1623
SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;

Query Match 70.3%; Score 104; DB 1; Length 1707;
Best Local Similarity 68.0%; Pred. No. 1.5e-06;
Matches 17; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 TMPEFMFCNNVNCNFAASNDYSYWL 25
Db 1538 TMPEFLYCNPCDVCYASRNDKSYWL 1562
|||||:||||:|||||:|||||
|||||:||||:|||||:|||||

RESULT 14
CA24 HUMAN
ID CA24 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2 (IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
collagen and comparison with the alpha 1(IV) chain.";
RL J. Biol. Chem. 263:19488-19493(1988).
[2]
RP SEQUENCE OF 1-1042 FROM N.A.
RC TISSUE=Placenta;

MEDLINE=88151998; PubMed=3345760;
Brazel D., Pollner R., Oberbaumer I., Kuehn K.;
"Human basement membrane collagen (type IV). The amino acid sequence
of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
chain reveals deletions in the alpha 1(IV) chain.";
Eur. J. Biochem. 172:35-42(1988).
[3]
RP SEQUENCE OF 1254-1712 FROM N.A.
RX MEDLINE=87219158; PubMed=3582677;
Hostikka S.L., Kurkinen M., Tryggvason K.;
"Nucleotide sequence coding for the human type IV collagen alpha 2
chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
(IV) but not with the collagenous domain or 3'-untranslated region.";
FEBS Lett. 216:281-286(1987).
[4]
RP SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87024438; PubMed=3025878;
Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
"Human collagen genes encoding basement membrane alpha 1 (IV) and
alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
[5]
RP SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
"Duplication of type IV collagen COOH-terminal repeats and species-
specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
J. Biol. Chem. 262:9231-9238(1987).
[6]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
"The structural genes for alpha 1 and alpha 2 chains of human type IV
collagen are divergently encoded on opposite DNA strands and have an
overlapping promoter region.";
J. Biol. Chem. 263:17217-17220(1988).
[7]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89030632; PubMed=2846280;
Poeschl E., Pollner R., Kuehn K.;
"The genes for the alpha 1(IV) and alpha 2(IV) chains of human
basement membrane collagen type IV are arranged head-to-head and
separated by a bidirectional promoter of unique structure.";
EMBO J. 7:2697-2695(1988).
[8]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=93305049; PubMed=8317999;
Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
"Identification of a novel sequence element in the common promoter
region of human collagen type IV genes, involved in the regulation of
divergent transcription.";
Biochem. J. 292:687-695(1993).
[9]
RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
Siebold B., Deutzmann R., Kuehn K.;
"The arrangement of intra- and intermolecular disulfide bonds in the
carboxyterminal, non-collagenous aggregation and cross-linking domain
of basement membrane type IV collagen.";
Eur. J. Biochem. 176:617-624(1988).
-!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
```



```
CC CC triple-helical 7S domain.
CC CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC CC these, located in the NCI domain, are conserved in all known type
CC CC IV collagens.
CC CC -----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X05562; CAA29076.1; -
CC CC EMBL; X05610; CAA29098.1; -
CC CC EMBL; J02760; AAA58422.1; -
CC CC EMBL; M36963; AAA53099.1; -
CC CC EMBL; X12784; CAA31225.1; -
CC CC EMBL; J04217; AAA53097.1; -
CC CC PIR; A32024; CGHU2B.
CC CC Genew; HGNC:2203; COL4A2.
CC CC MIM; 120090; -
CC CC GO; GO:000587; C:collagen type IV; TAS.
CC CC GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
CC CC GO; GO:0030198; P:extracellular matrix organization and bioge. . .; NAS.
CC CC InterPro; IPR008161; Clg_helix.
CC CC InterPro; IPR008160; Collagen.
CC CC Pfam; PF01413; C4; 2.
CC CC Pfam; PF01391; Collagen; 24.
CC CC ProDom; PD000007; Clg_helix; 7.
CC CC ProDom; PD003923; ProcollagnC4; 1.
CC CC SMART; SM00111; C4; 2.
CC CC DR; DR:GO00587; C:collagen type IV; TAS.
CC CC DR; DR:GO005201; F:extracellular matrix structural constituent; TAS.
CC CC DR; DR:GO0030198; P:extracellular matrix organization and bioge. . .; NAS.
CC CC DR; DR:InterPro; IPR008161; Clg_helix.
CC CC DR; DR:InterPro; IPR008160; Collagen.
CC CC DR; DR:Pfam; PF01413; C4; 2.
CC CC DR; DR:Pfam; PF01391; Collagen; 24.
CC CC DR; DR:ProDom; PD000007; Clg_helix; 7.
CC CC DR; DR:ProDom; PD003923; ProcollagnC4; 1.
CC CC DR; DR:SMART; SM00111; C4; 2.
CC CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC KW Glycoprotein; Basement membrane; Collagen; Signal.
CC FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
CC FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.
CC FT CHAIN 184 1712 TRIPLE-HELICAL REGION.
CC FT DOMAIN 184 1484 TRIPLE-HELICAL REGION (NCI).
CC FT DOMAIN 1485 1712 NONHELICAL REGION (NCI).
CC FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
CC FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
CC FT DISULFID 1549 1555 BY SIMILARITY.
CC FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
CC FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
CC FT DISULFID 1658 1665 BY SIMILARITY.
CC FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
CC FT CONFLICT 471 471 R -> P (IN REF. 2).
CC FT CONFLICT 683 683 A -> G (IN REF. 2).
CC FT CONFLICT 1575 1575 M -> I (IN REF. 5).
CC FT CONFLICT 1663 1663 G -> H (IN REF. 9).
CC FT CONFLICT 1701 1701 H -> G (IN REF. 9).
CC FT CONFLICT 1712 1712 AA; 167535 MW; 2582A17847890037 CRC64;
CC FT SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
CC Query Match 70.3%; Score 104; DB 1; Length 1712;
CC Best Local Similarity 68.0%; Pred. No. 1.5e-06;
CC Matches 17; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
CC QY 1 TMAPFCNNNNVCFASRNDYSYWL 25
CC DB 1543 TMAPFCNPGDVYASRNDKSYWL 1567
CC RESULT 15
CC CA44_BOVIN
CC ID CA44_BOVIN STANDARD; PRT; 453 AA.
CC AC Q29442;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
```

```
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Iens; PubMed=1370461;
RX MEDLINE=92112769; PubMed=1370461;
RA Mariyam M., Kalluri R., Hudson B.G., Readers S.T.;
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
RT collagens."
RT J. Biol. Chem. 267:1253-1258(1992).
RL [2]
RN SEQUENCE OF 217-246.
RP MEDLINE=90202779; PubMed=2318922;
RX Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RA "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen."
RT J. Biol. Chem. 265:5466-5469(1990).
RL [3]
RN SEQUENCE OF 217-233.
RP MEDLINE=87222419; PubMed=2438283;
RX Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen."
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77480; AAA30458.2; ALT_SEQ.
CC PIR; S18804; S18804.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagn4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 4.
CC ProDom; PD003923; ProcollagnC4; 1.
CC SMART; SM00111; C4; 2.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.
```

FT NON_TER 1 1
 FT DOMAIN <1 222 TRIPLE-HELICAL REGION.
 FT DISULFID 223 453 NONHELICAL REGION (NC1).
 FT DISULFID 243 332 OR 329 (BY SIMILARITY).
 FT DISULFID 276 329 OR 332 (BY SIMILARITY).
 FT DISULFID 288 294 BY SIMILARITY.
 FT DISULFID 351 449 OR 446 (BY SIMILARITY).
 FT DISULFID 385 446 OR 449 (BY SIMILARITY).
 FT DISULFID 397 404 BY SIMILARITY.
 FT CONFLICT 219 219 I -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A65BC1 CRC64;

Query Match 69.6%; Score 103; DB 1; Length 453;
 Best Local Similarity 64.0%; Pred No. 5, 5e-07;
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TMPFMFCINNVNCFASRNDYSYL 25
 DB 282 TLPFAYCNIHQVYARRNDRSYWL 306

Search completed: April 5, 2004, 06:59:38
 Job time : 3.1477 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 15.0121 Seconds
(without alignments)
525.440 Million cell updates/sec

Title: US-10-032-221b-38
Perfect score: 148
Sequence: 1 TMPEFNCINNVCFASRNDYSYWL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp Unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 148 | 100.0 | 179 | 11 P70165 | P70165 mus musculus |
| 2 | 148 | 100.0 | 253 | 11 Q61436 | Q61436 mus musculus |
| 3 | 148 | 100.0 | 585 | 11 Q80V57 | Q80V57 mus musculus |
| 4 | 148 | 100.0 | 799 | 11 Q8NS7 | Q8NS7 mus musculus |
| 5 | 148 | 100.0 | 886 | 4 Q9NUE7 | Q9NUE7 homo sapien |
| 6 | 148 | 100.0 | 1684 | 6 Q8HYC1 | Q8HYC1 canis famill |
| 7 | 148 | 100.0 | 1688 | 6 Q86622 | Q86622 canis famill |
| 8 | 148 | 100.0 | 1691 | 11 Q9ESQ2 | Q9ESQ2 mus musculus |
| 9 | 145 | 98.0 | 161 | 11 Q61430 | Q61430 mus musculus |
| 10 | 145 | 98.0 | 210 | 6 Q28273 | Q28273 canis famill |
| 11 | 145 | 98.0 | 225 | 6 Q28271 | Q28271 canis famill |
| 12 | 145 | 98.0 | 226 | 11 Q9SLQ8 | Q9SLQ8 mus musculus |
| 13 | 145 | 98.0 | 229 | 4 Q8NF88 | Q8NF88 homo sapien |
| 14 | 145 | 98.0 | 229 | 4 Q9NYC5 | Q9NYC5 homo sapien |
| 15 | 145 | 98.0 | 246 | 11 Q61435 | Q61435 mus musculus |
| 16 | 145 | 98.0 | 979 | 13 Q919K3 | Q919K3 gallus gall |

| | | | | | |
|----|-----|------|------|-----------|---------------------|
| 17 | 145 | 98.0 | 1075 | 4 Q86X41 | Q86X41 homo sapien |
| 18 | 145 | 98.0 | 1621 | 4 Q9HAR9 | Q9HAR9 homo sapien |
| 19 | 145 | 98.0 | 1669 | 11 Q9QZS0 | Q9QZS0 mus musculus |
| 20 | 140 | 94.6 | 203 | 6 Q29032 | Q29032 sus scrofa |
| 21 | 140 | 94.6 | 203 | 6 Q28682 | Q28682 cryetolagus |
| 22 | 140 | 94.6 | 212 | 6 Q28567 | Q28567 ovis aries |
| 23 | 139 | 93.9 | 212 | 6 Q28512 | Q28512 macaca mula |
| 24 | 139 | 93.9 | 230 | 11 Q63122 | Q63122 rattus norv |
| 25 | 139 | 93.9 | 245 | 4 Q9NYC4 | Q9NYC4 homo sapien |
| 26 | 139 | 93.9 | 1752 | 5 Q07265 | Q07265 strongyloce |
| 27 | 129 | 87.2 | 1747 | 5 Q26640 | Q26640 strongyloce |
| 28 | 121 | 81.8 | 1802 | 5 Q17163 | Q17163 brugia mala |
| 29 | 113 | 76.4 | 205 | 6 Q28274 | Q28274 canis famill |
| 30 | 113 | 76.4 | 546 | 11 Q99K97 | Q99K97 mus musculu |
| 31 | 113 | 76.4 | 1600 | 4 Q9UEH6 | Q9UEH6 homo sapien |
| 32 | 113 | 76.4 | 1691 | 11 Q9ESQ1 | Q9ESQ1 mus musculu |
| 33 | 108 | 73.0 | 1723 | 5 Q9GQB1 | Q9GQB1 hydra atten |
| 34 | 104 | 70.3 | 202 | 6 Q28272 | Q28272 canis famill |
| 35 | 104 | 70.3 | 312 | 11 Q64457 | Q64457 mus musculu |
| 36 | 104 | 70.3 | 358 | 11 Q91VI3 | Q91VI3 mus musculu |
| 37 | 104 | 70.3 | 673 | 4 Q14052 | Q14052 homo sapien |
| 38 | 104 | 70.3 | 1682 | 11 Q9QZ89 | Q9QZ89 mus musculu |
| 39 | 103 | 69.6 | 208 | 6 Q29468 | Q29468 canis famill |
| 40 | 96 | 64.9 | 713 | 5 Q9GV24 | Q9GV24 sarcophaga |
| 41 | 96 | 64.9 | 1024 | 5 Q8T7S4 | Q8T7S4 anopheles g |
| 42 | 96 | 64.9 | 1761 | 5 Q18407 | Q18407 drosophila |
| 43 | 96 | 64.9 | 1940 | 5 Q9VMV5 | Q9VMV5 drosophila |
| 44 | 95 | 64.2 | 1779 | 5 Q9VMV4 | Q9VMV4 drosophila |
| 45 | 69 | 46.6 | 854 | 5 Q09238 | Q09238 pseudocorti |

ALIGNMENTS

RESULT 1

ID P70165 PRELIMINARY; PRT; 179 AA.
AC P70165;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbaumer I.;
RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and alpha5(IV) in differentiated teratocarcinoma cells."
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82218; CAA57698.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER 1
FT NON_TER 179 179
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 100.0%; Score 148; DB 11; Length 179;
Best Local Similarity 100.0%; Pred. No. 2.3e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPEFNCINNVCFASRNDYSYWL 25

Db 36 TMPEFNCINNVCFASRNDYSYWL 60

```
RESULT 2
Q61436
ID Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sares J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
laminae: Sequence distribution, association with laminins, and
developmental switches.";
RT J. Cell Biol. 127:875-891 (1994).
RL EMBL; 235168; CAA84531.1; -.
DR PIR; I48304; I48304.
DR MGD; MGI:88456; Col14a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match 100.0%; Score 148; DB 11; Length 253;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPEMFCNNVNCNFSARNDSYWL 25
Db 83 TMPEMFCNNVNCNFSARNDSYWL 107

RESULT 3
Q80V57
ID Q80V57 PRELIMINARY; PRT; 585 AA.
AC Q80V57;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Col4a5 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J.J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
```

```
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043317; AAH43317.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
SQ SEQUENCE 585 AA; 59283 MW; 26774FE364F7FD8D CRC64;

Query Match 100.0%; Score 148; DB 11; Length 585;
Best Local Similarity 100.0%; Pred. No. 7.7e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPEMFCNNVNCNFSARNDSYWL 25
Db 415 TMPEMFCNNVNCNFSARNDSYWL 439

RESULT 4
Q8NS7
ID Q8NS7 PRELIMINARY; PRT; 799 AA.
AC Q8NS7;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGD; MGI:88456; Col14a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 100.0%; Score 148; DB 11; Length 799;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
DB 629 TMPFMFCNNVNCNPFASNDYSYWL 653

RESULT 5
Q9NUB7 PRELIMINARY; PRT; 886 AA.
ID Q9NUB7
AC Q9NUB7
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
DE COL4A5.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 100.0%; Score 148; DB 4; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
DB 716 TMPFMFCNNVNCNPFASNDYSYWL 740

RESULT 6
Q8HYC1 PRELIMINARY; PRT; 1684 AA.
ID Q8HYC1
AC Q8HYC1
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type IV collagen alpha 5 chain (Fragment).
DE COL4A5.
GN Canis familiaris (Dog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Harvey S.J., Zheng K., Jefferson B., Sado Y., Naito I., Ninomiya Y.,
RA Jacobs R., Thorner P.S.;
RL "Recombinant alpha5(IV) collagen: In vivo adenoviral-mediated gene
transfer to smooth muscle restores expression of the alpha6(IV)
collagen chain in a canine model of Alport syndrome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY078501; AAL83712.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.

QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
DB 1521 TMPFMFCNNVNCNPFASNDYSYWL 1545

RESULT 7
Q866Z2 PRELIMINARY; PRT; 1688 AA.
ID Q866Z2
AC Q866Z2
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type IV collagen alpha 5.
DE COL4A5.
GN Canis familiaris (Dog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Cox M.L., Lees G.E., Kashtan C.E., Murphy K.E.;
RL "Genetic Cause of X-linked Alport Syndrome in a Family of Domestic
Dogs."
RL Mamm. Genome 0:0-0(2003).
DR EMBL; AF470624; AAC33458.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
RP SEQUENCE FROM N.A.
RA Harvey S.J., Zheng K., Jefferson B., Sado Y., Naito I., Ninomiya Y.,
RA Jacobs R., Thorner P.S.;
RL "Recombinant alpha5(IV) collagen: In vivo adenoviral-mediated gene
transfer to smooth muscle restores expression of the alpha6(IV)
collagen chain in a canine model of Alport syndrome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY078501; AAL83712.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.

QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
DB 1521 TMPFMFCNNVNCNPFASNDYSYWL 1545

RESULT 8
Q9ESQ2 PRELIMINARY; PRT; 1691 AA.
ID Q9ESQ2
AC Q9ESQ2
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type IV collagen alpha 5 chain.
DE COL4A5.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20536494; PubMed=10965041;
RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,
RT Kishiro Y., Sado Y., Yoshioka H., Ninomiya Y.;
RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in
RT Epithelial Basement Membranes.";
RL J. Biochem. 128:427-434(2000).
DR EMBL; AB041350; BAB13673.1; -.
DR GO; GOI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Clg_helix.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match 100.0%; Score 148; DB 11; Length 1691;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25
DB 1521 TMPFMFCNNVNCNFCASRNDYSYWL 1545

RESULT 9
Q61430 PRELIMINARY; PRT; 161 AA.
ID Q61430;
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX Ohteraume I.;
RT "Cloning of the NCI domains to the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49489.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

Query Match 98.0%; Score 145; DB 11; Length 161;
Best Local Similarity 96.0%; Pred. No. 5.9e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25

```

```

DB 8 TMPFLFCNNVNCNFCASRNDYSYWL 32

RESULT 10
Q28273 PRELIMINARY; PRT; 210 AA.
ID Q28273;
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RX Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA823633D CRC64;

Query Match 98.0%; Score 145; DB 6; Length 210;
Best Local Similarity 96.0%; Pred. No. 7.7e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25
DB 51 TMPFLFCNNVNCNFCASRNDYSYWL 75

RESULT 11
Q28271 PRELIMINARY; PRT; 225 AA.
ID Q28271;
AC Q28271;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RX Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
RN [2]
RP SEQUENCE FROM N.A.

```

```
RC STRAIN-Sanoyed;
RA Thorne P.S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50933; AAC4583.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER 1 225
FT NON_TER 225
SQ SEQUENCE 225 AA; 24585 MW; 2C20455850416E47 CRC64;

Query Match 98.0%; Score 145; DB 6; Length 225;
Best Local Similarity 96.0%; Pred. No. 8.2e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 65 TMPFLFCNNVNCVFASRNDYSYWL 89

RESULT 12
Q99LQ8 PRELIMINARY; PRT; 226 AA.
AC Q99LQ8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; SC002269; RAH02269.1; -.
DR MGD; MGT:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Hypothetical Protein.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 226 AA; 25042 MW; 4F7FD5371181C21 CRC64;

Query Match 98.0%; Score 145; DB 11; Length 226;
Best Local Similarity 96.0%; Pred. No. 8.3e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 56 TMPFLFCNNVNCVFASRNDYSYWL 80

RESULT 13
Q8NF88 PRELIMINARY; PRT; 229 AA.
AC Q8NF88;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN SEQUENCE FROM N.A.
RP He A.B.;
RT "Cloning and Expression of Arresten in Escherichia coli and Pachia pastoris.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF536207; AAM97359.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 229 AA; 25391 MW; 09B21FDSAB517E9E CRC64;

Query Match 98.0%; Score 145; DB 4; Length 229;
Best Local Similarity 96.0%; Pred. No. 8.4e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 59 TMPFLFCNNVNCVFASRNDYSYWL 83

RESULT 14
Q9NYC5 PRELIMINARY; PRT; 229 AA.
AC Q9NYC5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H., Takahashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K., Erickson M.B., Dhanabal M., Simons M., Post M., Kufe D., Weichselbaum R.R., Sukhatme V.P., Kalluri R.;
RT "Anti-angiogenic cues from vascular basement membrane collagen.";
RL Cancer Res. 0:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Fu J., Bai X., Wang W., Ruan C.;
RT "Arresten, a collagen-derived inhibitor of angiogenesis.";
RL Chung Hua Heueh Yeh Heueh Tea Chih 22:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Peng X., Yin B., Yuan J., Qiang B.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Zheng Q.C., Song Z.F., Zheng Y.W., Li Y.Q., Shu X.;
RT "Molecular cloning and sequencing of human arresten gene.";
RL Zhonghua Shi Yan Wai Ke Za Zhi 19:46-47(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Song Z.F., Zheng Q.C.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258349; AAF72630.1; -.
DR EMBL; AF363672; AAK53382.1; -.
DR EMBL; AF400431; AAK92480.1; -.
DR EMBL; AY285780; AAP43112.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
```

FT NON TER 1 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5C1D5 CRC64;

Query Match 98.0%; Score 145; DB 4; Length 229;
Best Local Similarity 96.0%; Pred. No. 8.4e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TMPFMFCNINNVCFASRNDYSYWL 25
DB 59 TMPFMFCNINNVCFASRNDYSYWL 83

RESULT 15

O61435 PRELIMINARY; PRT; 246 AA.
AC O61435;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
CN COL4A3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H.; Sane J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: Sequence, distribution, association with laminins, and
RT developmental switches".
RL J. Cell Biol. 127:879-891(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z35166; CAA84529.1; -.
DR PIR; I48302; I48302.
DR MGI; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
FT NON-TER 1
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 98.0%; Score 145; DB 11; Length 246;
Best Local Similarity 96.0%; Pred. No. 9e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TMPFMFCNINNVCFASRNDYSYWL 25
DB 75 TMPFMFCNINNVCFASRNDYSYWL 99

Search completed: April 5, 2004, 07:03:57
Job time: 15.0121 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 22.5182 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221b-38

Perfect score: 148

Sequence: 1 TWFFMFCNNVNCNFASTRNDYSYWL 25.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 148 | 100.0 | 25 | 6 | ADA20237 T7 mutant |
| 2 | 148 | 100.0 | 229 | 7 | ADC17699 Human typ |
| 3 | 148 | 100.0 | 284 | 2 | AAY31995 Type IV c |
| 4 | 148 | 100.0 | 264 | 3 | AAY37557 Human alp |
| 5 | 148 | 100.0 | 309 | 3 | AAB54044 Human pan |
| 6 | 148 | 100.0 | 772 | 2 | AAR23873 Human alp |
| 7 | 148 | 100.0 | 772 | 2 | AAR23873 Human alp |
| 8 | 148 | 100.0 | 1685 | 4 | AAB09643 Human typ |
| 9 | 148 | 100.0 | 1693 | 4 | ABG04839 Novel hum |
| 10 | 145 | 98.0 | 229 | 1 | AAP35524 Complete |
| 11 | 145 | 98.0 | 229 | 3 | AAY67943 Human typ |
| 12 | 145 | 98.0 | 229 | 5 | AAY75587 Human typ |
| 13 | 145 | 98.0 | 229 | 6 | ADA20217 Human typ |
| 14 | 145 | 98.0 | 229 | 7 | ADC17695 Human typ |
| 15 | 145 | 98.0 | 260 | 2 | AAY31991 Type IV c |
| 16 | 145 | 98.0 | 260 | 3 | AAY37553 Human alp |
| 17 | 145 | 98.0 | 406 | 3 | AAB58169 lung can |
| 18 | 145 | 98.0 | 1669 | 4 | AAM40863 Human pol |
| 19 | 145 | 98.0 | 1669 | 5 | ABB90760 Human Tum |
| 20 | 145 | 98.0 | 1659 | 5 | ABB57334 Mouse isc |
| 21 | 145 | 98.0 | 1659 | 6 | ABU54467 Human tum |
| 22 | 145 | 98.0 | 1672 | 4 | AAM39077 Human pol |
| 23 | 140 | 94.6 | 471 | 2 | AAR79163 Partial s |
| 24 | 140 | 94.6 | 471 | 2 | AAY44171 Bovine ty |
| 25 | 140 | 94.6 | 471 | 3 | AAY56783 Bovine al |

| | | | | | | |
|----|-----|------|-----|---|----------|--------------------|
| 26 | 140 | 94.6 | 471 | 4 | AAE09483 | Aae09483 Bovine al |
| 27 | 139 | 93.9 | 25 | 6 | ADA20236 | Ada20236 T7 peptid |
| 28 | 139 | 93.9 | 79 | 5 | AAY75600 | Aay75600 Human typ |
| 29 | 139 | 93.9 | 79 | 6 | ADA20264 | Ada20264 Human tum |
| 30 | 139 | 93.9 | 88 | 5 | AAU75608 | Aau75608 Human typ |
| 31 | 139 | 93.9 | 88 | 5 | AAU75607 | Aau75607 Human typ |
| 32 | 139 | 93.9 | 88 | 6 | ADA20271 | Ada20271 Human tum |
| 33 | 139 | 93.9 | 88 | 6 | ADA20272 | Ada20272 Human tum |
| 34 | 139 | 93.9 | 124 | 5 | AAU75594 | Aau75594 Human typ |
| 35 | 139 | 93.9 | 124 | 6 | ADA20258 | Ada20258 Human tum |
| 36 | 139 | 93.9 | 132 | 5 | AAU75597 | Aau75597 Human typ |
| 37 | 139 | 93.9 | 132 | 6 | ADA20261 | Ada20261 Human tum |
| 38 | 139 | 93.9 | 191 | 5 | AAU75596 | Aau75596 Human typ |
| 39 | 139 | 93.9 | 191 | 6 | ADA20260 | Ada20260 Human tum |
| 40 | 139 | 93.9 | 211 | 3 | AAY95918 | Aay95918 Human Goo |
| 41 | 139 | 93.9 | 211 | 5 | ABG79208 | Abg79208 Human GP |
| 42 | 139 | 93.9 | 218 | 2 | AAR79164 | Aar79164 Partial s |
| 43 | 139 | 93.9 | 218 | 2 | AAY44172 | Aay44172 Human typ |
| 44 | 139 | 93.9 | 218 | 3 | AAY56784 | Aay56784 Human alp |
| 45 | 139 | 93.9 | 218 | 4 | AAE09484 | Aae09484 Human alp |

ALIGNMENTS

RESULT 1

ADA20237

ID ADA20237 standard; peptide; 25 AA.

XX AC ADA20237;

XX DT 20-NOV-2003 (first entry)

XX DE T7 mutant peptide related to human type IV collagen and angiogenesis.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX KW metastasis; basement membrane organisation; type IV collagen network;

XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;

XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX KW cytosolic; gene therapy; T7 mutant peptide; mutant; mutatin;

XX KW type IV collagen alpha 3 chain; tumatatin; human.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 5 /note= "Wild-type Leu substituted by Met"

FT Misc-difference 9 /note= "Wild-type Val substituted by Ile"

FT Misc-difference 11 /note= "Wild-type Asp substituted by Asn"

XX WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX XX WPI; 2003-587256/55.

XX DR New peptide, useful for preparing a composition for inhibiting tumor

XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX XX Claim 61; Page 45; 240pp; English.

XX PS This invention relates to novel isolated proteins and their fragments

XX CC

CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the mutated T7 peptide of the
 CC invention. The wild-type T7 peptide sequence is given in Seq ID37 (see
 CC ADA20236) and this was derived from the amino acid sequence of tumstatin,
 CC which in turn was derived from the amino acid sequence of human type IV
 CC collagen alpha 3 chain.

XX Sequence 25 AA;

Query Match 100.0%; Score 148; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSRNDYSYWL 25

Db 1 TMPFMFCNNVNCNFSRNDYSYWL 25

RESULT 2

ADCL7699

XX ADC17699 standard; protein; 229 AA.

XX ADC17699;

XX 18-DEC-2003 (first entry)

XX Human type IV collagen alpha 5 chain protein SEQ ID NO:306.

XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

XX Homo sapiens.

XX W02003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

XX 29-OCT-2001; 2001US-0351289P.

XX 22-MAR-2002; 2002US-0366854P.

XX 03-JUN-2002; 2002US-0385362P.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

XX (SUND/) SUNDARAMOORTHY M.

XX (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.

XX Disclosure; SEQ ID NO 306; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (7) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (8) identifying inhibitors of type IV
 CC domain hexamer of type IV collagen; (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
 CC anticancer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents an amino acid sequence which is used in the exemplification of
 CC the present invention.

XX Sequence 229 AA;

Query Match 100.0%; Score 148; DB 7; Length 229;

Best Local Similarity 100.0%; Pred. No. 1.6e-13;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSRNDYSYWL 25

Db 59 TMPFMFCNNVNCNFSRNDYSYWL 83

RESULT 3

AAV31995

XX AAV31995 standard; protein; 264 AA.

XX AAV31995;

XX 05-JAN-2000 (first entry)

XX Type IV collagen NC1 domain alpha-5 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
 KW rheumatoid arthritis; retinal neovascularization;
 KW choroidal neovascularization; macular degeneration;
 KW corneal neovascularization; retinopathy of prematurity;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidemic keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; soggren's; acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW ulcer; herpes simplex infection; Herpes zoster infection;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; polyarteritis; Wegener's sarcooidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;
 KW sarcooid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Bales disease; Bechets disease; myopia;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;

KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
KW fibrovascular tissue proliferation; haemangiomas; Osler-Weber-Rendu; AIDS;
KW ocular neovascular disease; osteoarthritis; chronic inflammation;
KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
KW pemphigoid.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT Peptide 1..17
FT Protein /note= "BM40 signal peptide"
FT Protein 18..264
FT Peptide /note= "mature protein"
FT Peptide 18..25
FT Protein /note= "affinity tag"
FT Protein 26..264
FT Protein /note= "NC1 alpha-5 monomer"
XX
XX WC9949885-A2.
XX
XX
XX 07-OCT-1999.
XX 26-MAR-1999; 99WC-US006445.
XX
XX 27-MAR-1998; 98US-0079783P.
PR 29-OCT-1998; 98US-0106170P.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX
PI Hudson BG, Sarraz MP;
XX
XX WPI; 1999-601297/51.
DR N-PSDB; AA220093.
XX
XX Inhibition of angiogenesis with non-collagenous alpha chain monomer
PT useful for treating e.g. tumor growth or metastasis, neovascularisation,
PT etc.
XX
XX
XX Disclosure; Fig 17e; 56pp; English.
XX
XX This sequence represents a recombinant type IV collagen non-collagenous
CC (NC1) domain alpha-5 polypeptide composed of a BM40 signal sequence
CC (which is cleaved from the mature protein) to facilitate protein
CC secretion, and a mature protein comprising an affinity tag (facilitates
CC purification and identification of the material) and the alpha-1 chain
CC monomer. The invention provides methods and kits for inhibiting
CC angiogenesis, tumor growth and metastasis, and endothelial cell
CC interaction with the extracellular matrix, each method comprising
CC contacting the tumour or animal tissue with 1 or more isolated type IV
CC collagen NC1 alpha chain monomer(s) selected from the group consisting of
CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AA931991-
CC 96). The monomers can be produced via recombinant protein expression. The
CC polynucleotides and polypeptides are used to treat an angiogenesis-
CC mediated disorder or condition, especially selected from solid and blood-
CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
CC neovascularization, choroidal neovascularization, macular degeneration,
CC corneal neovascularization, retinopathy of prematurity, corneal graft
CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
CC keratitis, superior limbic keratitis, pterygium keratitis sicca, sogrens,
CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
CC simplex infections, herpes zoster infections, protozoan infections,
CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
CC keratolysis, trauma, systemic lupus, polyarthritis, Wegener's
CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,
CC sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,
CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Behcet's
CC disease, myopia, optic pits, Stargards disease, pars planitis, chronic
CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
CC complications, abnormal proliferation of fibrovascular tissue.

CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)
XX
SQ Sequence 264 AA;

Query Match 100.0%; Score 148; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMCNINNVNCFASRNDYSYWL 25
Db 94 TMPFMCNINNVNCFASRNDYSYWL 118

RESULT 4

AA937557
ID AAY97557 standard; protein; 264 AA.

XX
AC AAY97557;

XX
DT 12-FEB-2001 (first entry)

XX
DE Human alpha5(IV)NC1 protein sequence.

XX
KW Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
KW tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
KW chronic inflammation; psoriasis; therapy; alpha5(IV)NC1.

OS Homo sapiens.

XX
PN WO2000059532-A1.

XX
PD 12-OCT-2000.

XX
PF 31-MAR-2000; 2000WO-US008678.

XX
PR 01-APR-1999; 99US-0127391P.

XX
PA (BIOS-) BIOSTRATUM INC.

XX
PI Brooks P, Hudson B;

XX
DR WPI; 2000-664962/64.

XX
N-PSDB; AAA90995.

XX
PT Use of antagonists of specific integrin receptors for inhibiting
PT angiogenesis, tumor growth or metastases, or endothelial cell
PT interactions with the extracellular matrix.

PS Disclosure; Fig 17e; 78pp; English.

XX This sequence is a human type IV collagen alpha chain monomer, designated
CC alpha5(IV)NC1. The invention relates to a method for inhibiting
CC angiogenesis, tumor growth or metastases, or endothelial cell
CC interactions with the extracellular matrix, comprising contacting the
CC cells or tissue with a polypeptide composition containing antagonists of
CC specific integrin receptors. The methods and the antagonists are useful
CC for inhibiting angiogenesis, tumor growth or metastases, or endothelial
CC cell interaction with the extracellular matrix. The antagonists are also
CC useful for treating diseases and conditions with accompanying undesired
CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
CC treating non-tumorigenic diseases and conditions with accompanying
CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
CC retinal neovascularisation, choroidal neovascularisation, muscular
CC degeneration, corneal graft rejection, vitamin A deficiency, atopic

CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
CC sickle cell anaemia, sarcoïd, carotid obstructive disease, post-laser
CC complications, chronic inflammation or psoriasis
XX
SQ Sequence 264 AA;

Query Match 100.0%; Score 148; DB 3; Length 264;
Best Local Similarity 100.0%; Pred No. 1.9e-13; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

QY 1 TMPEMFCNNVNCNPFASRNDYSYWL 25
|||||
Db 94 TMPEMFCNNVNCNPFASRNDYSYWL 118
|||||

RESULT 5
AAB54044
ID AAB54044 standard; protein; 309 AA.

XX AAB54044;

XX 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:496.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
KW diagnosis; identification; cytostatic; neuroprotective; nootropic;
KW immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiact; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic; neural;
KW immune system; muscular; reproductive; gastrointestinal; pulmonary;
KW cardiovascular; renal; proliferative.

XX Homo sapiens.

XX W0200055320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005989.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-579444/54.

XX N-PSDB; AAC98809.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition.

XX Claim 11; Page 934-935; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiact and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including

CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention
XX

SQ Sequence 309 AA;

Query Match 100.0%; Score 148; DB 3; Length 309;
Best Local Similarity 100.0%; Pred. No. 2.2e-13; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

QY 1 TMPEMFCNNVNCNPFASRNDYSYWL 25
|||||
Db 139 TMPEMFCNNVNCNPFASRNDYSYWL 163
|||||

RESULT 6

AAR23873

ID AAR23873 standard; protein; 772 AA.

XX AAR23873;

XX 25-NOV-1992 (first entry)

XX Human alpha 5 (IV) of type IV collagen.

XX Mutations; Alport's syndrome; basement membranes; diabetes mellitus.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 43.47 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 159.160 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 275.277 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 334.336 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 456.458 /note= "interruption in Gly-X-Y sequence"

XX US5114840-A.

XX 19-MAY-1992.

XX 07-JUL-1989; 89US-00377238.

XX 07-JUL-1989; 89US-00377238.

XX (TRYG/) TRYGGVASON K.

XX Tryggvason K, Hostikka SL;

XX WPI; 1992-192174/23.

XX N-PSDB; AAQ24551.

XX Isolation of DNA encoding alpha-5(IV) polypeptide of type IV collagen - to
PT detect mutations in genes for alpha-5(IV) chain which produce genetic or
PT acquired basement membrane disorders e.g. Alport's syndrome.
XX Disclosure; Fig 2; 14pp; English.

XX The sequence is that of the alpha 5(IV) polypeptide chain of human type
CC IV collagen, the major component of basement membranes. The protein
CC contains the Gly-X-Y repeat coding sequence typical for collagenous
CC proteins at one end and a typical NC-domain coding sequence at the other
CC end. The sequence can be used to detect mutations in individual genes
CC specific for this chain which can, directly or indirectly, produce
CC several human diseases. It can also be used to determine genetic, e.g.
CC Alport's syndrome, or acquired e.g. diabetes mellitus, disorders of the
CC basement membrane, and as probes or antibodies against these nucleotide

CC sequences. Gene fragments generated through amplifications from human
CC genomic or cloned DNA can also be used for detection and analysis of
CC genes
XX
SQ Sequence 772 AA;
Query Match 100.0%; Score 148; DB 2; Length 772;
Best Local Similarity 100.0%; Pred. No. 5.9e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TMPFMFCNNVNCNPFASNDYSYWL 25
Db 602 TMPFMFCNNVNCNPFASNDYSYWL 626
RESULT 7
AAW09643
ID AAW09643 standard; protein; 772 AA.
XX
AC AAW09643;
XX
DT 25-MAR-2003 (revised)
DT 16-JUN-1997 (first entry)
XX
DE Human type IV collagen alpha-5.
XX
KW Collagen alpha5(IV); basement membrane; Alport's syndrome; nephritis;
KW kidney; renal failure; antibody; diagnosis; COL4A5 gene; X chromosome.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..543
FT /label= Collagenous domain
FT /note= "collagenous" domain contains Gly-X-Y tripeptide
FT repeats, interrupted at positions 43-47, 159-160, 275-
FT 276, 334-335, 456-459"
FT 544..772
FT /label= Non-collagenous_domain
FT 742..751
FT /label= Immunogenic peptide
FT /note= "peptide used to raise diagnostic antibodies
FT (claim 1)"
XX
XX US5593900-A.
XX
XX 14-JAN-1997.
XX
XX 11-OCT-1994; 54US-00321084.
XX
XX 07-JUL-1989; 89US-00377238.
XX 20-DEC-1990; 90US-00630563.
XX
XX (TRYG/) TRYGGVASON K.
XX (HOST/) HOSTIKKA S L.
XX (HOYH/) HOYHTYA M.
XX
XX Hostikka SL, Tryggvason K, Hoyhtya M;
XX
XX WPI; 1997-099481/09.
XX N-PSDB; AAT47812.
XX
XX New antibodies specific for human type IV collagen alpha5 chain - used to
XX detect absence of this chain in patients with renal failure.
XX
XX Disclosure; Fig 2A-2B; 12pp; English.
XX
XX The amino acid sequence of a portion (AAW09643) of the previously known
XX human type IV collagen chain, alpha5(IV), was deduced from cDNA clones
XX (see also AAT47812) obt'd. using probes based on conserved sequences of
XX human alpha1(IV) and alpha2(IV) collagen chains and of the Drosophila
XX alpha(IV) chain. It includes a complete non-collagenous domain that shows
XX 83% identity with that of alpha1(IV) and 63% with that of the alpha2(IV)

CC chain. Mutations in the alpha5(IV) gene (COL4A5) are associated with
CC Alport's syndrome. Antibodies raised against a peptide (see also
CC AAW09644) specific to alpha5(IV) can be used in the diagnosis of basement
CC membrane disorders such as Alport's syndrome. (Updated on 25-MAR-2003 to
CC correct pf field.)
XX
SQ Sequence 772 AA;
Query Match 100.0%; Score 148; DB 2; Length 772;
Best Local Similarity 100.0%; Pred. No. 5.9e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TMPFMFCNNVNCNPFASNDYSYWL 25
Db 602 TMPFMFCNNVNCNPFASNDYSYWL 626
RESULT 8
ABG04839
ID ABG04839 standard; protein; 1685 AA.
XX
AC ABG04839;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4830.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS69026.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 35198; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1685 AA;
Query Match 100.0%; Score 148; DB 4; Length 1685;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMPFMFCNNVNCNPFASRNDYSYWL 25
DB 1515 TMPFMFCNNVNCNPFASRNDYSYWL 1539
RESULT 9
ABG15619
ID ABG15619 standard; protein; 1693 AA.
XX
AC ABG15619;
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15610.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR
DR N-PSDB; AAS79806.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 45978; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. NOTE: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1693 AA;
Query Match 100.0%; Score 148; DB 4; Length 1693;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMPFMFCNNVNCNPFASRNDYSYWL 25
DB 1523 TMPFMFCNNVNCNPFASRNDYSYWL 1547
RESULT 10
AAP93524
ID AAP93524 standard; protein; 229 AA.
XX
AC AAP93524;
XX
DT 25-MAR-2003 (revised)
DT 03-OCT-2002 (revised)
DT 04-JUN-1990 (first entry)
XX
DE Complete sequence of the alpha-1-NCl domain of type IV collagen.
XX
DE Alpha-1-NCl domain; type IV collagen; cell adhesion; heparin;
KW aortic endothelial cells; metastatic carcinoma M4 cells; rat fibroblasts;
KW MM fibrosarcoma cells; C6 glioma cell; A431 breast carcinoma cells;
KW wound healing; implant acceptance.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 17..27
FT /note= "TS-3"
FT Peptide 49..60
FT /note= "TS-2"
FT Peptide 201..216
FT /note= "TS-1"
XX
FN WO8903392-A.
XX
XX 20-APR-1989.
PD
XX 20-AUG-1988; 88WO-US003023.
PF
XX 08-OCT-1987; 87US-00106858.
PR
XX (MINU) MINNESOTA UNIVERSITY.
PA
XX Tsilbary EC;
PI
XX WPI; 1989-130015/17.
DR
XX Polypeptide(s) with type IV collagen activity - used to promote wound
PT healing, implant acceptance and cellular attachment and inhibit malignant
PT cells.
XX
PS Fig 2; page 1/12; 40pp; English.
XX
CC The peptides in the features table are claimed (Claim 1, p. 22). They
CC were synthesised using the Merrifield solid phase method. Binding assays
CC were carried out using peptides TS-1, TS-2 and TS-3. TS-1 promotes
CC adhesions of aortic endothelial cells, metastatic carcinoma M4 cells,
CC normal rat fibroblasts, MM fibrosarcoma cells, C6 glioma cells and A431
CC breast carcinoma cells. TS-2 binds to type IV collagen, to heparin and
CC promotes adhesion of the above cells. Peptides TS-1, TS-2 and TS-3 may be
CC attached to promote wound healing and implant acceptance, promote cellular
CC attachment to culture substrata or inhibit the metastasis of malignant
CC cells. They may be used to coat a prosthetic device. (Updated on 03-OCT-
CC 2002 to add missing OS field.) (Updated on 25-MAR-2003 to correct PF
CC field.) (Updated on 25-MAR-2003 to correct PI field.)
XX

SQL Sequence 229 AA;

Query Match 98.0%; Score 145; DB 1; Length 229;
Best Local Similarity 96.0%; Pred. No. 4.4e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 59 TMPFLFCNNVNCVFASRNDYSYWL 83

RESULT 11

AAV67943
ID AAY67943 standard; protein; 229 AA.

XX AC AAY67943;

XX DT 03-APR-2000 (first entry)

XX DE Human type IV collagen alpha 1 chain protein sequence SEQ ID NO:2.

XX KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
ocular angiogenesis disease; Osler-Weber Syndrome; telangiectasia;
myocardial angiogenesis; plaque neovascularisation; angiofibroma;
atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
contraception; obesity.

XX OS Homo sapiens.

XX PN WO9965940-A1.

XX PD 23-DEC-1999.

XX PF 17-JUN-1999; 99WO-US013737.

XX PR 17-JUN-1998; 98US-0089689P.

XX PR 25-MAR-1999; 99US-0126175P.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2000-097708/08.

XX DR N-PSDB; AAZ571159.

XX PT Anti-angiogenic proteins comprising the NCI domain of the alpha 1, 2 or 3
chain of Type IV collagen used in, e.g. treatment of benign tumors and
rheumatoid arthritis.

XX PS Example 1; Fig 1B; 117pp; English.

XX CC The present sequence represents the human type IV collagen alpha 1 chain.
The present invention describes an isolated protein chosen from the NCI
domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a
fragment, analogue, derivative or mutant, which has anti-angiogenic
properties. The anti-angiogenic proteins, multimers and chimeras are
useful for inhibiting angiogenic activity in mammalian tissue, especially
for treating diseases chosen from angiogenesis-dependent cancers, benign
tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular
angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,
plaque neovascularisation, telangiectasia, haemophilic joints,
angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,
scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori
ulcers, dialysis graft vascular access stenosis, contraception and
obesity. The compositions can be used to inhibit a disease characterised
by angiogenic activity, in conjunction with radiation therapy,
chemotherapy or immunotherapy

XX SQL Sequence 229 AA;

Query Match 98.0%; Score 145; DB 3; Length 229;
Best Local Similarity 96.0%; Pred. No. 4.4e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 59 TMPFLFCNNVNCVFASRNDYSYWL 83

RESULT 12

AAU75587
ID AAU75587 standard; protein; 229 AA.

XX AC AAU75587;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 1 chain.

XX KW Human; type IV collagen alpha 1 chain; cytostatic; antiangiogenic;
non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
endothelial cell proliferation; apoptosis; Arresten; Canstatin;
Tumstatin; angiogenesis; tumour.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2002-188037/24.

XX DR N-PSDB; ABK15359.

XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
treating disorders involving angiogenesis.

XX PS Example 1; Fig 1B; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
domain, having one or more of the characteristics selected from: (a) the
ability to bind alphavbeta3 integrin; (b) the ability to inhibit
proliferation of endothelial cells; and (c) the ability to cause
apoptosis of endothelial cells. Also described are the following: (1) use
of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
analogue or allelic variant in the preparation of a medicament for
treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
where the angiogenesis is mediated by one or more endothelial cell
integrins or one or more endothelial cell integrin subunits; or (b) by
promoting or inducing endothelial cell apoptosis in a tissue, where the
endothelial cell apoptosis is mediated by one or more endothelial cell
integrins or one or more endothelial cell integrin subunits; (2) use of
an antibody or peptide that specifically binds the alpha1, alpha2,
alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the
preparation of a medicament for inhibiting angiogenesis or cell
proliferation; (3) use of an inhibitor, such as an antibody, antibody
fragment or peptide of receptor-mediated angiogenesis in the preparation
of a medicament for treating a proliferative disease in a vertebrate,
where the disease is characterised by angiogenesis that is mediated by
receptors to Arresten, Canstatin or Tumstatin and where the receptors
inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
the presence of a medicament for promoting angiogenesis in a tissue; and
(5) use of integrins in the preparation of a medicament for promoting or
inducing angiogenesis or cell proliferation in a tissue. The fragments
Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 1 chain
XX
SQ Sequence 229 AA;

Query Match 98.0%; Score 145; DB 5; Length 229;

Best Local Similarity 96.0%; Pred. No. 4.4e-13; Indels 0; Gaps 0;
Matches 24; Conservative 1; Mismatches 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25
Db 59 TMPFLFCNNVNCNFCASRNDYSYWL 83

RESULT 13

ADA20217
ID ADA20217 standard; protein; 229 AA.

XX AC ADA20217;

XX AC 20-NOV-2003 (first entry)
XX DE Human type IV collagen alpha 1 chain partial protein sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytoskeletal; gene therapy; alpha 1 chain; arresten; human.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587255/55.
XX DR N-PSDB; ADA20216.

XX PT New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 101; Fig 1; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is the partial amino acid sequence of the alpha 1 chain of human
CC type IV collagen. The "arresten" peptide of the invention was derived
CC from this protein.

XX SQ Sequence 229 AA;

Query Match 98.0%; Score 145; DB 6; Length 229;

Best Local Similarity 96.0%; Pred. No. 4.4e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASRNDYSYWL 25

Db 59 TMPFLFCNNVNCNFCASRNDYSYWL 83

RESULT 14

ADCI7695
ID ADCI7695 standard; protein; 229 AA.

XX AC ADCI7695;

XX DT 18-DEC-2003 (first entry)

XX DE Human type IV collagen alpha 1 chain protein SEQ ID NO:302.

XX crystallised NCI domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
XX tumour metastasis inhibitor; tumour growth inhibitor;
XX endothelial cell interaction inhibitor;
XX basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
XX anti-angiogenic; ophthalmological; antiarteriosclerotic; antiulcer;
XX endothelial cell adhesion inhibitor;
XX endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
XX ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX blood-borne tumour.

XX OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX PA (UNIV) UNIV KANSAS MEDICAL CENT.
XX PA (SUND/) SUNDARAMOORTHY M.

XX PA (HUDS/) HUDSON B.

XX PI Sundaramoorthy M, Hudson B;

XX DR WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.

XX Disclosure; SEQ ID NO 302; 168pp; English.

XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina

CC membrane formation in cell or tissue development; (7) a crystal of an NCI
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
 CC antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and
 CC anticancer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents an amino acid sequence which is used in the exemplification of
 CC the present invention.
 XX Sequence 229 AA;

Query Match 98.0%; Score 145; DB 7; Length 229;
 Best Local Similarity 96.0%; Pred. No. 4.4e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25

DB 59 TMPFLFCNNVNCVFASRNDYSYWL 83

RESULT 15

AAV31991

ID AAY31991 standard; protein; 260 AA.

XX AAY31991;

XX 05-JAN-2000 (first entry)

DE Type IV collagen NCI domain alpha-1 monomer.

KW Type IV collagen; NCI domain; non-collagenous domain; human;
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
 KW rheumatoid arthritis; retinal neovascularization;
 KW choroidal neovascularization; macular degeneration;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidemic keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; soggrens; acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW ulcer; Herpes simplex infection; Herpes zoster infection;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; polyarteritis; Wegener's sarcooidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;
 KW sarcooid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu;
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
 KW pemphigoid.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..17

FT /note= "BM40 signal peptide"

FT 18..260

FT /note= "mature protein"

FT Peptide 18..25
 FT /note= "affinity tag"
 FT Protein 26..260
 FT /note= "NCI alpha-1 monomer"

PN WO9949885-A2.

XX 07-OCT-1999.

XX 26-MAR-1999; 99WO-US006445.

XX 27-MAR-1998; 98US-0079783P.

XX 29-OCT-1998; 98US-0106170P.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Sarraz MP;

XX WPI; 1999-601297/51.

XX N-PSDB; AAZ20089.

PT Inhibition of angiogenesis with non-collagenous alpha chain monomer
 useful for treating e.g. tumor growth or metastasis, neovascularisation,
 PT etc.

XX Disclosure; Fig 17a; 56pp; English.

XX This sequence represents a recombinant type IV collagen non-collagenous
 CC (NCI) domain alpha-1 polypeptide composed of a BM40 signal sequence
 CC (which is cleaved from the mature protein) to facilitate protein
 CC secretion, and a mature protein comprising an affinity tag (facilitates
 CC purification and identification of the material) and the alpha-1 chain
 CC monomer. The invention provides methods and kits for inhibiting
 CC angiogenesis, tumour growth and metastasis, and endothelial cell
 CC interaction with the extracellular matrix, each method comprising
 CC contacting the tumour or animal tissue with 1 or more isolated type IV
 CC collagen NCI alpha chain monomer(s) selected from the group consisting of
 CC alpha-1, alpha-2, alpha-3 and alpha-6 NCI chain monomers (see AAY31991-
 CC 96). The monomers can be produced via recombinant protein expression. The
 CC polynucleotides and polypeptides are used to treat an angiogenesis-
 CC mediated disorder or condition, especially selected from solid and blood-
 CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
 CC neovascularization, choroidal neovascularization, macular degeneration,
 CC corneal neovascularization, retinopathy of prematurity, corneal graft
 CC rejection, neovascular glaucoma, A deficiency, contact lens overwear, atopic
 CC keratoconjunctivitis, vitamin A deficiency, retrolental fibroplasia, epidemic
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, soggrens,
 CC acne rosacea, phlyctenulosis, syphilis, Mycobacteria infections, lipid
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
 CC simplex infections, herpes zoster infections, protozoan infections,
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
 CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's
 CC sarcooidosis, scleritis, Steven's Johnson disease, radial keratotomy,
 CC sickle cell anaemia, sarcooid, pseudoxanthoma elasticum, Pagets disease,
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
 CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
 CC complications, abnormal proliferation of fibrovascular tissue,
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)

XX Sequence 260 AA;

Query Match 98.0%; Score 145; DB 2; Length 260;

Best Local Similarity 96.0%; Pred. No. 5e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25

DB 90 TMPFLFCNNVNCVFASRNDYSYWL 114

Search completed: April 5, 2004, 06:58:31
Job time : 22.5182 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 15.678 Seconds

(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-38

Perfect score: 148

Sequence: 1 TMSFMCNINNVCFASRNDYSYL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 148 | 100.0 | 229 | 14 | US-10-032-221B-38 |
| 2 | 148 | 100.0 | 229 | 14 | US-10-032-221B-38 |
| 3 | 148 | 100.0 | 309 | 9 | US-09-925-297-496 |
| 4 | 145 | 98.0 | 229 | 14 | US-10-032-221B-38 |
| 5 | 145 | 98.0 | 229 | 14 | US-10-032-221B-38 |
| 6 | 145 | 98.0 | 406 | 9 | US-09-925-302-507 |
| 7 | 145 | 98.0 | 1669 | 15 | US-10-372-683-8 |
| 8 | 139 | 93.9 | 25 | 14 | US-10-032-221B-37 |
| 9 | 139 | 93.9 | 79 | 14 | US-10-032-221B-36 |
| 10 | 139 | 93.9 | 88 | 14 | US-10-032-221B-33 |
| 11 | 139 | 93.9 | 88 | 14 | US-10-032-221B-33 |
| 12 | 139 | 93.9 | 124 | 14 | US-10-032-221B-20 |
| 13 | 139 | 93.9 | 132 | 14 | US-10-032-221B-23 |
| 14 | 139 | 93.9 | 191 | 14 | US-10-032-221B-22 |
| 15 | 139 | 93.9 | 211 | 14 | US-10-270-877-46 |

Sequence 46, Appl
Sequence 304, Appl
Sequence 10, Appl
Sequence 48095, A
Sequence 7032, Ap
Sequence 5832, Ap
Sequence 39, Appl
Sequence 267, Appl
Sequence 38021, A
Sequence 307, Appl
Sequence 265, Appl
Sequence 266, Appl
Sequence 261, Appl
Sequence 303, Appl
Sequence 305, Appl
Sequence 518, Appl
Sequence 27, Appl
Sequence 30, Appl
Sequence 9, Appl
Sequence 259, Appl
Sequence 40, Appl
Sequence 260, Appl
Sequence 255, Appl
Sequence 42, Appl
Sequence 253, Appl
Sequence 41, Appl
Sequence 230, Appl
Sequence 210, Appl
Sequence 270, Appl

16 139 93.9 211 14 US-10-270-837-46
17 139 93.9 232 14 US-10-206-699-304
18 139 93.9 244 14 US-10-032-221B-10
19 138 93.2 46 9 US-09-864-761-48095
20 124 83.8 1759 15 US-10-369-493-7032
21 121 81.8 1744 15 US-10-369-493-5832
22 117 79.1 27 14 US-10-032-221B-39
23 115 77.7 22 14 US-10-206-699-267
24 113 76.4 142 9 US-09-864-761-38021
25 113 76.4 228 14 US-10-206-699-307
26 112 75.7 22 14 US-10-206-699-265
27 106 71.6 22 14 US-10-206-699-266
28 105 70.9 18 14 US-10-206-699-261
29 104 70.3 227 14 US-10-206-699-303
30 104 70.3 227 14 US-10-032-221B-6
31 104 70.3 231 14 US-10-206-699-305
32 104 70.3 430 9 US-09-925-302-518
33 104 70.3 459 15 US-10-331-496A-27
34 104 70.3 459 15 US-10-372-683-30
35 104 70.3 1712 10 US-09-961-403-9
36 102 68.9 18 14 US-10-206-699-259
37 97 65.5 27 14 US-10-032-221B-40
38 96 64.9 18 14 US-10-206-699-260
39 94 63.5 18 14 US-10-206-699-255
40 93 62.8 27 14 US-10-032-221B-42
41 91 61.5 18 14 US-10-206-699-253
42 90 60.8 19 14 US-10-032-221B-41
43 90 60.8 20 14 US-10-206-699-290
44 89 60.1 15 14 US-10-206-699-210
45 89 60.1 22 14 US-10-206-699-270

ALIGNMENTS

RESULT 1

US-10-032-221B-38
; Sequence 38, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghubram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (Formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1995-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T7-mutant (amino acids 73-97 of SEQ ID NO:10; methionine has be
; OTHER INFORMATION: substituted for the leucine residue at position 77 of the full-
; OTHER INFORMATION: enyth Tumstatin molecule, and isoleucine has been substituted f
; OTHER INFORMATION: valine at position 81, and asparagine has been substituted for
; OTHER INFORMATION: spartic acid at position 83)
US-10-032-221B-38

Query Match 100.0%; Score 148; DB 14; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2.2e-14;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25
 DB 1 TMEFMCNINNVCFASRNDYSYWL 25

RESULT 2
 US-10-206-699-306
 ; Sequence 306, Application US/10206699
 ; Publication No. US20030100510A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sundaramoorthy, M.
 ; APPLICANT: Hudson, B.
 ; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
 ; FILE REFERENCE: MBHB 01-1017
 ; CURRENT APPLICATION NUMBER: US/10/206,699
 ; CURRENT FILING DATE: 2002-07-26
 ; PRIOR APPLICATION NUMBER: US 60/308,523
 ; PRIOR FILING DATE: 2001-07-27
 ; PRIOR APPLICATION NUMBER: US 60/351,289
 ; PRIOR FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: US 60/366,854
 ; PRIOR FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: US 60/385,362
 ; PRIOR FILING DATE: 2002-06-03
 ; NUMBER OF SEQ ID NOS: 307
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 306
 ; LENGTH: 229
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: alpha 5 chain
 US-10-206-699-306

Query Match 100.0%; Score 148; DB 14; Length 229;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25
 DB 59 TMEFMCNINNVCFASRNDYSYWL 83

RESULT 3
 US-09-925-297-496
 ; Sequence 496, Application US/09925297
 ; Patent No. US20020081659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA105
 ; CURRENT APPLICATION NUMBER: US/09/925,297
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05989
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 928
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 496
 ; LENGTH: 309
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (247)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-297-496

Query Match 100.0%; Score 148; DB 9; Length 309;
 Best Local Similarity 100.0%; Pred. No. 2.5e-13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25
 DB 139 TMEFMCNINNVCFASRNDYSYWL 163

RESULT 4
 US-10-206-699-302
 ; Sequence 302, Application US/10206699
 ; Publication No. US20030100510A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sundaramoorthy, M.
 ; APPLICANT: Hudson, B.
 ; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
 ; FILE REFERENCE: MBHB 01-1017
 ; CURRENT APPLICATION NUMBER: US/10/206,699
 ; CURRENT FILING DATE: 2002-07-26
 ; PRIOR APPLICATION NUMBER: US 60/308,523
 ; PRIOR FILING DATE: 2001-07-27
 ; PRIOR APPLICATION NUMBER: US 60/351,289
 ; PRIOR FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: US 60/366,854
 ; PRIOR FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: US 60/385,362
 ; PRIOR FILING DATE: 2002-06-03
 ; NUMBER OF SEQ ID NOS: 307
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 302
 ; LENGTH: 229
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: alpha 1 chain
 US-10-206-699-302

Query Match 98.0%; Score 145; DB 14; Length 229;
 Best Local Similarity 96.0%; Pred. No. 5e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25
 DB 59 TMEFMCNINNVCFASRNDYSYWL 83

RESULT 5
 US-10-032-221B-2
 ; Sequence 2, Application US/10032221B
 ; Publication No. US20030144481A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kalluri, Raghuram
 ; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
 ; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
 ; CURRENT APPLICATION NUMBER: US/10/032,221B
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: PCT/US01/00565
 ; PRIOR FILING DATE: 2001-01-08
 ; PRIOR APPLICATION NUMBER: US 09/625,191
 ; PRIOR FILING DATE: 2000-07-21
 ; PRIOR APPLICATION NUMBER: US 09/543,371
 ; PRIOR FILING DATE: 2000-04-04
 ; PRIOR APPLICATION NUMBER: US 09/479,118
 ; PRIOR FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: US 09/335,224
 ; PRIOR FILING DATE: 1999-06-17
 ; PRIOR APPLICATION NUMBER: US 60/126,175
 ; PRIOR FILING DATE: 1999-03-25
 ; PRIOR APPLICATION NUMBER: US 60/089,689
 ; PRIOR FILING DATE: 1998-06-17

```

; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-2

Query Match      98.0%; Score 145; DB 14; Length 229;
Best Local Similarity 96.0%; Pred. No. 5e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25
Db 59 TWPFFMFCNNVNCNPFASRNDYSYWL 83

RESULT 6
US-09-925-302-507
; Sequence 507, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT FILING DATE: 2001-08-10
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 507
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; FEATURE:
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-507

Query Match      98.0%; Score 145; DB 9; Length 406;
Best Local Similarity 96.0%; Pred. No. 8.7e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25
Db 236 TWPFFMFCNNVNCNPFASRNDYSYWL 260

RESULT 7
US-10-372-683-8
; Sequence 8, Application US/10372683
; Publication No. US20040009171A1
; GENERAL INFORMATION:
; APPLICANT: GERRITSEN, MARY E.
; APPLICANT: PEALE JR., FRANKLIN V.
; APPLICANT: WU, THOMAS D.
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
; FILE REFERENCE: P1928R1P1
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 10/372,683
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/344,534
; PRIOR FILING DATE: 2001-10-18
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 8
; LENGTH: 1669
; TYPE: PRT
; ORGANISM: Homo sapien
```

```

US-10-372-683-8

Query Match      98.0%; Score 145; DB 15; Length 1669;
Best Local Similarity 96.0%; Pred. No. 3.4e-12;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25
Db 1499 TWPFFMFCNNVNCNPFASRNDYSYWL 1523

RESULT 8
US-10-032-221B-37
; Sequence 37, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREC
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T7 (amino acids 73-97 of SEQ ID NO:10)
US-10-032-221B-37

Query Match      93.9%; Score 139; DB 14; Length 25;
Best Local Similarity 88.0%; Pred. No. 4.4e-13;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25
Db 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25

RESULT 9
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREC
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
```

PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/089,689
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 26
LENGTH: 79
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26

Query Match 93.9%; Score 139; DB 14; Length 79;
Best Local Similarity 88.0%; Pred. No. 1.3e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPPMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
DB 20 TMPPFLFCNVNVCVFASRNDYSYWL 44

RESULT 10
US-10-032-221B-33
Sequence 33, Application US/10032221B
Publication No. US20030144481A1
GENERAL INFORMATION:
APPLICANT: Kalluri, Raghuram
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/335,224
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/089,689
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33
LENGTH: 88
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)
US-10-032-221B-33

Query Match 93.9%; Score 139; DB 14; Length 88;
Best Local Similarity 88.0%; Pred. No. 1.5e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPPMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
DB 29 TMPPFLFCNVNVCVFASRNDYSYWL 53

RESULT 11
US-10-032-221B-34
Sequence 34, Application US/10032221B
Publication No. US20030144481A1
GENERAL INFORMATION:
APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/335,224
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/089,689
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 34
LENGTH: 88
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine has been substituted for the cysteine residue at position 125 of the full-length Tumstatin molecule)
OTHER INFORMATION: the full-length Tumstatin molecule)
US-10-032-221B-34

Query Match 93.9%; Score 139; DB 14; Length 88;
Best Local Similarity 88.0%; Pred. No. 1.5e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPPMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
DB 29 TMPPFLFCNVNVCVFASRNDYSYWL 53

RESULT 12
US-10-032-221B-20
Sequence 20, Application US/10032221B
Publication No. US20030144481A1
GENERAL INFORMATION:
APPLICANT: Kalluri, Raghuram
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/335,224
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/089,689
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 124
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match 93.9%; Score 139; DB 14; Length 124;
Best Local Similarity 88.0%; Pred. No. 2e-12; 0; Indels 0; Gaps 0;
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
Db 73 TMPFLFCNVNDVCFASRNDYSYWL 97

RESULT 13

US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match 93.9%; Score 139; DB 14; Length 132;
Best Local Similarity 88.0%; Pred. No. 2.2e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
Db 73 TMPFLFCNVNDVCFASRNDYSYWL 97

RESULT 14

US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175

; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match 93.9%; Score 139; DB 14; Length 191;
Best Local Similarity 88.0%; Pred. No. 3.1e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
Db 20 TMPFLFCNVNDVCFASRNDYSYWL 44

RESULT 15

US-10-270-877-46
; Sequence 46, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match 93.9%; Score 139; DB 14; Length 211;
Best Local Similarity 88.0%; Pred. No. 3.4e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
|||:||||:|||||
Db 73 TMPFLFCNVNDVCFASRNDYSYWL 97

Search completed: April 5, 2004, 07:36:06
Job time : 15.678 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 5.87167 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221B-38
Perfect score: 148
Sequence: 1 TWPFFMFCNNVCFASRNDYSYWL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 148 | 100.0 | 264 | 4 | US-09-589-927-10 |
| 2 | 148 | 100.0 | 264 | 4 | US-09-277-665-10 |
| 3 | 148 | 100.0 | 264 | 4 | US-09-589-987-10 |
| 4 | 145 | 98.0 | 260 | 4 | US-09-589-927-2 |
| 5 | 145 | 98.0 | 260 | 4 | US-09-277-665-2 |
| 6 | 145 | 98.0 | 260 | 4 | US-09-589-987-2 |
| 7 | 140 | 94.6 | 471 | 2 | US-08-399-889-24 |
| 8 | 140 | 94.6 | 471 | 3 | US-09-167-364-24 |
| 9 | 140 | 94.6 | 471 | 3 | US-09-439-997-2 |
| 10 | 139 | 93.9 | 211 | 4 | US-09-512-563C-46 |
| 11 | 139 | 93.9 | 218 | 3 | US-08-399-889-25 |
| 12 | 139 | 93.9 | 218 | 3 | US-09-167-364-25 |
| 13 | 139 | 93.9 | 218 | 3 | US-09-439-997-4 |
| 14 | 139 | 93.9 | 268 | 4 | US-09-589-927-6 |
| 15 | 139 | 93.9 | 268 | 4 | US-09-277-665-6 |
| 16 | 139 | 93.9 | 268 | 4 | US-09-589-987-6 |
| 17 | 113 | 76.4 | 260 | 4 | US-09-589-927-12 |
| 18 | 113 | 76.4 | 260 | 4 | US-09-277-665-12 |
| 19 | 113 | 76.4 | 260 | 4 | US-09-589-987-12 |
| 20 | 104 | 70.3 | 258 | 4 | US-09-589-927-4 |
| 21 | 104 | 70.3 | 258 | 4 | US-09-277-665-4 |
| 22 | 104 | 70.3 | 258 | 4 | US-09-589-987-4 |
| 23 | 104 | 70.3 | 260 | 4 | US-09-589-927-8 |
| 24 | 104 | 70.3 | 260 | 4 | US-09-277-665-8 |
| 25 | 104 | 70.3 | 260 | 4 | US-09-589-987-8 |
| 26 | 89 | 60.1 | 1694 | 1 | US-08-494-168-2 |
| 27 | 46 | 31.1 | 83 | 4 | US-09-543-681A-8079 |

| | | | | | | |
|----|------|------|------|---|---------------------|--------------------|
| 28 | 46 | 31.1 | 160 | 4 | US-09-568-673B-7 | Sequence 7, Appli |
| 29 | 45.5 | 30.7 | 704 | 3 | US-08-792-832A-2 | Sequence 2, Appli |
| 30 | 45 | 30.4 | 867 | 4 | US-09-668-673B-2 | Sequence 2, Appli |
| 31 | 45 | 30.4 | 871 | 4 | US-09-773-426A-3 | Sequence 3, Appli |
| 32 | 45 | 30.4 | 1611 | 4 | US-09-668-673B-16 | Sequence 16, Appli |
| 33 | 44 | 29.7 | 49 | 1 | US-07-865-166A-6 | Sequence 6, Appli |
| 34 | 44 | 29.7 | 359 | 1 | US-08-307-382-2 | Sequence 2, Appli |
| 35 | 44 | 29.7 | 359 | 1 | US-08-366-779-2 | Sequence 2, Appli |
| 36 | 44 | 29.7 | 359 | 1 | US-08-478-727-2 | Sequence 2, Appli |
| 37 | 44 | 29.7 | 359 | 1 | US-08-473-508-2 | Sequence 2, Appli |
| 38 | 44 | 29.7 | 359 | 1 | US-08-789-936-2 | Sequence 2, Appli |
| 39 | 44 | 29.7 | 359 | 2 | US-08-833-610-6 | Sequence 6, Appli |
| 40 | 44 | 29.7 | 359 | 3 | US-08-834-033A-16 | Sequence 16, Appli |
| 41 | 44 | 29.7 | 359 | 4 | US-08-934-254-2 | Sequence 2, Appli |
| 42 | 44 | 29.7 | 359 | 4 | US-09-377-452-6 | Sequence 6, Appli |
| 43 | 44 | 29.7 | 359 | 4 | US-09-665-775-2 | Sequence 2, Appli |
| 44 | 44 | 29.7 | 575 | 4 | US-09-107-532A-4554 | Sequence 4554, Ap |
| 45 | 43.5 | 29.4 | 354 | 4 | US-09-574-942-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-09-589-927-10
; Sequence 10, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-10

Query Match 100.0%; Score 148; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVCFASRNDYSYWL 25
DB 94 TWPFFMFCNNVCFASRNDYSYWL 118

RESULT 2
US-09-277-665-10
; Sequence 10, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-10

Query Match 100.0%; Score 148; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
QY 1 TMPFMFCNNVNCNFCASNDYSYWL 25
DB 94 TMPFMFCNNVNCNFCASNDYSYWL 118

RESULT 3
US-09-589-987-10
; Sequence 10, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-10

Query Match 100.0%; Score 148; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASNDYSYWL 25
DB 94 TMPFMFCNNVNCNFCASNDYSYWL 118

RESULT 4
US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 98.0%; Score 145; DB 4; Length 260;
Best Local Similarity 96.0%; Pred. No. 9.6e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASNDYSYWL 25
DB 90 TMPFMFCNNVNCNFCASNDYSYWL 114

RESULT 5
US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-277-665-2

Query Match 94.6%; Score 140; DB 2; Length 471;
Best Local Similarity 92.0%; Pred. No. 8.8e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASNDYSYWL 25
DB 90 TMPFMFCNNVNCNFCASNDYSYWL 114

RESULT 6
US-09-589-987-2
; Sequence 2, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-2

Query Match 98.0%; Score 145; DB 4; Length 260;
Best Local Similarity 96.0%; Pred. No. 9.6e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASNDYSYWL 25
DB 90 TMPFMFCNNVNCNFCASNDYSYWL 114

RESULT 7
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match 94.6%; Score 140; DB 2; Length 471;
Best Local Similarity 92.0%; Pred. No. 8.8e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFCASNDYSYWL 25
DB 90 TMPFMFCNNVNCNFCASNDYSYWL 114
```

```
Db 300 TWPFLFCNVNVCNPFASNDYSYWL 324

RESULT 8
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match 94.6%; Score 140; DB 3; Length 471;
Best Local Similarity 92.0%; Pred. No. 8.8e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFMFCNINNVNCFASNDYSYWL 25
Db 300 TWPFLFCNVNVCNPFASNDYSYWL 324

RESULT 9
US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 94.6%; Score 140; DB 3; Length 471;
Best Local Similarity 92.0%; Pred. No. 8.8e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFMFCNINNVNCFASNDYSYWL 25
Db 300 TWPFLFCNVNVCNPFASNDYSYWL 324

RESULT 10
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
```

```
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 93.9%; Score 139; DB 4; Length 211;
Best Local Similarity 88.0%; Pred. No. 5.3e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFMFCNINNVNCFASNDYSYWL 25
Db 73 TWPFLFCNVNVCNPFASNDYSYWL 97

RESULT 11
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

Query Match 93.9%; Score 139; DB 2; Length 218;
Best Local Similarity 88.0%; Pred. No. 5.4e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFMFCNINNVNCFASNDYSYWL 25
Db 47 TWPFLFCNVNVCNPFASNDYSYWL 71

RESULT 12
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match 93.9%; Score 139; DB 3; Length 218;
Best Local Similarity 88.0%; Pred. No. 5.4e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFMFCNINNVNCFASNDYSYWL 25
Db 47 TWPFLFCNVNVCNPFASNDYSYWL 71
```

Best Local Similarity 88.0%; Pred. No. 5.4e-12; Indels 0; Gaps 0;
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCNNVNCNFAASNDYSYWL 25
Db 47 TMPFLFCNVNDVCNFAASNDYSYWL 71

RESULT 13

US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1283-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match 93.9%; Score 139; DB 3; Length 218;
Best Local Similarity 88.0%; Pred. No. 5.4e-12; Indels 0; Gaps 0;
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCNNVNCNFAASNDYSYWL 25
Db 47 TMPFLFCNVNDVCNFAASNDYSYWL 71

RESULT 14

US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match 93.9%; Score 139; DB 4; Length 268;
Best Local Similarity 88.0%; Pred. No. 6.7e-12; Indels 0; Gaps 0;
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCNNVNCNFAASNDYSYWL 25
Db 97 TMPFLFCNVNDVCNFAASNDYSYWL 121

RESULT 15

US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match 93.9%; Score 139; DB 4; Length 268;
Best Local Similarity 88.0%; Pred. No. 6.7e-12; Indels 0; Gaps 0;
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCNNVNCNFAASNDYSYWL 25
Db 97 TMPFLFCNVNDVCNFAASNDYSYWL 121

Search completed: April 5, 2004, 07:07:25
Job time : 6.87167 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.5569 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221B-39

Perfect score: 151

Sequence: 1 KQRTTMTPLFCNVNDVCFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 146 | 96.7 | 220 | B49736 | collagen alpha 3(I |
| 2 | 146 | 96.7 | 1670 | CGHU3B | collagen alpha 3(I |
| 3 | 145 | 96.0 | 471 | A39024 | collagen alpha 3(I |
| 4 | 140 | 92.7 | 161 | A49488 | collagen alpha 3(I |
| 5 | 140 | 92.7 | 246 | I48302 | collagen alpha 3(I |
| 6 | 130 | 88.1 | 253 | I48304 | collagen alpha 5(I |
| 7 | 130 | 86.1 | 754 | A35267 | collagen alpha 5(I |
| 8 | 130 | 86.1 | 1691 | I32917 | collagen alpha 5(I |
| 9 | 129 | 85.4 | 258 | B61228 | collagen alpha 1(I |
| 10 | 129 | 85.4 | 1669 | 1 CGHU4B | collagen alpha 1(I |
| 11 | 129 | 85.4 | 1669 | 1 CGMS4B | collagen alpha 1(I |
| 12 | 120 | 79.5 | 1747 | 2 A54121 | collagen alpha-4 c |
| 13 | 120 | 79.5 | 1752 | 2 A45407 | collagen alpha 3(I |
| 14 | 115 | 78.2 | 1758 | 2 T93350 | hypocretinal prote |
| 15 | 115 | 76.2 | 1759 | 2 T93351 | collagen alpha 2(I |
| 16 | 115 | 76.2 | 1763 | 2 S16366 | collagen alpha 2(I |
| 17 | 111 | 73.5 | 261 | 2 A34476 | collagen alpha 2(I |
| 18 | 107 | 70.9 | 1744 | 2 S40991 | collagen alpha 1(I |
| 19 | 105 | 69.5 | 1691 | 1 CGHU6B | collagen alpha 6(I |
| 20 | 102 | 67.5 | 1775 | 2 A61228 | collagen alpha 2(I |
| 21 | 102 | 67.5 | 1707 | 2 A33526 | collagen alpha 2(I |
| 22 | 102 | 67.5 | 1712 | 1 CGHU2B | collagen type IV a |
| 23 | 89 | 58.9 | 1761 | 2 T13990 | collagen alpha 4(I |
| 24 | 88 | 58.3 | 312 | 2 I48303 | collagen alpha 4(I |
| 25 | 88 | 58.3 | 623 | 2 A45137 | collagen alpha 4(I |
| 26 | 88 | 58.3 | 1690 | 1 CGHU1B | collagen alpha 4(I |
| 27 | 88 | 58.3 | 1775 | 2 A31893 | collagen alpha 1(I |
| 28 | 87 | 57.6 | 453 | 2 S18804 | collagen alpha 4(I |
| 29 | 62 | 41.1 | 79 | 2 C43928 | probable collagen |

RESULT 1
B49736
Collagen alpha 3(IV) chain, medium splice form - human (fragment)
N;Contains: collagen alpha 3(IV) chain, splice form GP-V
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text_change 17-Mar-2000
C;Accession: B49736; D49736; S69111
R;Peng, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A;Title: Alternative splicing of the NCI domain of the human alpha3(IV) collagen gene.
A;Reference number: A49736; MUID:94124597; PMID:8294492
A;Accession: B49736
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 169-220 <PEN1>
A;Accession: D49736
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 22-220 <FEN2>
A;Cross-references: GB:U02519; NID:9409106; PIDN:AAA18942.1; PID:9409107
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Carvera, J.; Wi
Eur. J. Biochem. 229, 754-760, 1995
A;Title: Characterization and expression of multiple alternatively spliced transcripts
utancigen and one of its alternative forms.
A;Reference number: S69111; MUID:95278230; PMID:7758473
A;Accession: S69111
A;Molecule type: mRNA
A;Residues: 1-45,169-204,'L',206-220 <PEN>
A;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.
C;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-2q37
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace
F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pred
F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status
F;22-220/Domain: carboxyl-terminal nonhelical, NCI <NCI>
F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 96.7%; Score 146; DB 2; Length 220;

Best Local Similarity 100.0%; Pred. No. 4.7e-14; Mismatches 0; Gaps 0;

Matches 26; Conservative 0; Indels 0;

QY 2 QRTTMTPLFCNVNDVCFASRNDYS 27

DB 78 QRTTMTPLFCNVNDVCFASRNDYS 103

RESULT 2

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human
N;Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 22-Jun-1999
C;Accession: A54763; A43928; A44043; A45971; A39786
R;Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A;Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
A;Reference number: A54763; MUID:94364994; PMID:8083201
A;Accession: A54763
A;Molecule type: mRNA
A;Residues: 1-1670 <VAR>
A;Cross-references: GB:X80031; NID:G577563; PID:G577564
A;Experimental source: kidney
R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha
A;Reference number: A43928; MUID:92147878; PMID:1737849
A;Accession: A43928
A;Molecule type: mRNA
A;Residues: 1331-1524, 'I', 1526-1670 <TUR>
A;Cross-references: GB:M81379
A;Experimental source: kidney
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A;Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
A;Reference number: A44043; MUID:93015826; PMID:1400291
A;Accession: A44043
A;Molecule type: DNA; mRNA
A;Residues: 1386-1670 <QUI>
A;Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896
A;Note: sequence extracted from NCBI backbone (NCBIP:115597)
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A;Reference number: A44738; MUID:94274734; PMID:8060044
A;Contents: annotation; erratum; correction to intronic sequence in A44043
R;Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A;Reference number: A54971; MUID:93280184; PMID:8505332
A;Accession: A54971
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1427-1444 <BER>
A;Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly ident
R;Morrison, K.E.; Mariyama, M.; Yang-Peng, T.L.; Reiders, S.T.
Am. J. Hum. Genet. 49, 545-554, 1997
A;Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
A;Reference number: A39786; MUID:91353570; PMID:1892840
A;Accession: A39786
A;Molecule type: mRNA
A;Residues: 1453-1593, 'A', 1595-1670 <MOR>
A;Cross-references: GB:S55790; NID:G234418; PIDN:AA319637.1; PID:G234419
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (ed and subsequently O-glycosylated.
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
C;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-q37
A;Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1595/3; 1643/2 #status incomplete
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-l
C;Function:
A;Description: minor structural component of extracellular basement membrane in kidney g
A;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
F;1-28/Domain: signal sequence #status predicted <STG>
F;29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <VAT>
F;29-42/Domain: amino-terminal nonhelical, NH1 <NH1>

F;43-1438/Region: interrupted helical
F;791-793/Region: cell attachment (R-G-D) motif
F;996-998/Region: cell attachment (R-G-D) motif
F;1154-1156/Region: cell attachment (R-G-D) motif
F;1306-1308/Region: cell attachment (R-G-D) motif
F;1345-1347/Region: cell attachment (R-G-D) motif
F;1432-1434/Region: cell attachment (R-G-D) motif
F;1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;31.33.39.41.125.422.476.479.682.722.809.1387/Disulfide bonds: interchain #status pred
F;253/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1460-1548.1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F;1505-1511.1616-1622/Disulfide bonds: #status predicted
F;1570-1662.1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
Query Match 96.7%; Score 146; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 3.3e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRFTHMPLFCNVNDVNCVFNASRNDYS 27
DB 1495 QRFTHMPLFCNVNDVNCVFNASRNDYS 1520
RESULT 3
A39024
collagen alpha 3(IV) chain - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1992 #sequence revision 04-Dec-1992 #text_change 13-Aug-1999
C;Accession: A39024; S20672; S17802; A35167; S39419; S13747; S20815
R;Morrison, K.E.; Germino, G.G.; Reiders, S.T.
J. Biol. Chem. 266, 34-39, 1991
A;Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
A;Reference number: A39024; MUID:91093146; PMID:1985905
A;Accession: A39024
A;Molecule type: mRNA
A;Residues: 1-471 <MOR>
A;Cross-references: EMBL:M63139; NID:G162886; PIDN:AAA2708.1; PID:G162887
R;Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J. Biol. Chem. 262, 7874-7877, 1987
A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
A;Reference number: S18432; MUID:87222419; PMID:2438283
A;Accession: S20672
A;Molecule type: protein
A;Residues: 227-228, 'X', 230-244 <BUT>
R;Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
J. Biol. Chem. 263, 13374-13380, 1988
A;Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
A;Reference number: S17802; MUID:88330844; PMID:3417661
A;Accession: S17802
A;Molecule type: protein
A;Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
R;Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J. Biol. Chem. 265, 5466-5469, 1990
A;Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type
A;Reference number: A35167; MUID:90202779; PMID:2318822
A;Accession: A35167
A;Molecule type: protein
A;Residues: 236-255 <GUN>
R;Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; No
J. Biol. Chem. 266, 15318-15324, 1991
A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol
A;Reference number: A39419; MUID:91332055; PMID:1869555
A;Accession: C39419
A;Molecule type: protein
A;Residues: 236-255 <GUN>
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;
F;1-238/Domain: collagenous (fragment) #status predicted <COL>
F;239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;239-353/Domain: repeat NC1 #status predicted <NC1>
F;354-471/Domain: repeat NC1 #status predicted <NC1>

F:232,238/Modified site: hydroxyproline (Pro) #status experimental
F:306-312,417-423/Disulfide bonds: #status predicted

Query Match 96.0%; Score 145; DB 2; Length 471;
Best Local Similarity 96.2%; Pred. No. 1.4e-13;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFCNVNDVNCVFASRNDYS 27
DB 296 QRFTTMPFLFCNVNDVNCVFASRNDYS 321

RESULT 4

S49488
collagen alpha 3(IV) chain - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C:Accession: S49488
R:Oberbaumer, I.

submitted to the EMBL Data Library, October 1994
A:Description: Cloning of the NC1 domains fo the minor collagen IV chains of mouse via B
ells

A:Reference number: S49487

A:Accession: S49488

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-161 <OB>

A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G5559916

C:Superfamily: collagen alpha 1(IV) chain

Query Match 92.7%; Score 140; DB 2; Length 161;

Best Local Similarity 92.3%; Pred. No. 2.7e-13;

Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFCNVNDVNCVFASRNDYS 27
DB 4 QRFTTMPFLFCNVNDVNCVFASRNDYS 29

RESULT 5

I48302
collagen alpha 3(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
C:Accession: I48302; S47278
R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48302

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-246 <RES>

A:Cross-references: EMBL:X35166; NID:G535197; PID:G535198

C:Superfamily: collagen alpha 1(IV) chain

Query Match 92.7%; Score 140; DB 2; Length 246;

Best Local Similarity 92.3%; Pred. No. 4e-13;

Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFCNVNDVNCVFASRNDYS 27
DB 71 QRFTTMPFLFCNVNDVNCVFASRNDYS 96

RESULT 6

I48304
collagen alpha 5(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999
C:Accession: I48304; S47280
R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48304

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-253 <RES>

A:Cross-references: EMBL:X35168; NID:G535201; PIDN:CAA84531.1; PID:G535202

C:Superfamily: collagen alpha 1(IV) chain

Query Match 86.1%; Score 130; DB 2; Length 253;

Best Local Similarity 80.8%; Pred. No. 1.2e-11;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFCNVNDVNCVFASRNDYS 27
DB 79 RRFSTMPFPCNVNDVNCVFASRNDYS 104

RESULT 7

A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267

R;Zheng, K.; Thorne, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994

A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-1

en type IV

A:Reference number: A55267; MUID:94224868; PMID:8171024

A:Accession: A55267

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>

A:Cross-references: GB:U07888; NID:G469547; PIDN:AA560258.1; PID:G469548

C:Superfamily: collagen alpha 1(IV) chain

Query Match 86.1%; Score 130; DB 2; Length 754;

Best Local Similarity 80.8%; Pred. No. 3.4e-11;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFCNVNDVNCVFASRNDYS 27
DB 587 RRFSTMPFPCNVNDVNCVFASRNDYS 612

RESULT 8

S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N:Alternate names: procollagen alpha 5(IV) chain
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A3

R;Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.

J. Biol. Chem. 267, 12475-12481, 1992

A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ident

n Alport syndrome patient.

A:Reference number: S22917; MUID:92316923; PMID:1352287

A:Accession: S22917

A:Molecule type: mRNA

A:Residues: 1-967 <ZHO>

A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234

R;Zhou, J.; Leinonen, A.; Tryggvason, K.

J. Biol. Chem. 269, 6608-6614, 1994

A:Title: Structure of the human type IV collagen COL4A5 gene.

A:Reference number: A54365; MUID:94165049; PMID:8120014

A:Accession: A54365

A:Molecule type: DNA

A:Residues: 1-922 <ZH2>

A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAAC27816.1; PI

R;Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paeppe, A.; Tryggva

Science 261, 1167-1169, 1993

A:Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited s

[illegible]

Db 84 RKFTMPFLFCNNVNCNPNFASNDYS 109

RESULT 10

CGHU4B

collagen alpha 1(IV) chain precursor - human
N;Alternate names: procollagen alpha 1(IV) chain

C;Species: Homo sapiens (man)

C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999

C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58

R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 264, 13565-13571, 1989

A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll

A;Reference number: S16876; MUID:89340433; PMID:2701944

A;Accession: S16876

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1669 <SO1>

A;Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAAS3098.1; PID:G180803

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988

R;Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988

A;Title: The structural genes for alpha1 and alpha2(IV) chains of human type IV collagen are

A;Reference number: A92690; MUID:89034231; PMID:3182844

A;Accession: A32117

A;Molecule type: DNA

A;Residues: 1-28 <SO12>

A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PID:G953233

R;Poeschl, E.; Pollner, R.; Kuehn, K.

EMBO J. 7, 2687-2695, 1988

A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c

A;Reference number: S02738; MUID:89030632; PMID:2846280

A;Accession: S02738

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-6, 'L', '8-28 <POE>

A;Cross-references: EMBL:X12784; NID:G30072

R;Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.

Eur. J. Biochem. 168, 599-636, 1987

A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem

A;Reference number: S00048; MUID:88029471; PMID:3311751

A;Accession: S00048

A;Molecule type: mRNA

A;Residues: 1-318, 'A', 320-944 <BRA1>

A;Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067

A;Accession: S25826

A;Molecule type: protein

A;Residues: 271-318, 'A', 320-554 <BRA2>

R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.

Eur. J. Biochem. 152, 213-219, 1985

A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S

A;Reference number: A23115; MUID:86004708; PMID:4043082

A;Accession: A23115

A;Molecule type: protein

A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

A;Experimental source: placenta

A;Note: the amino end of the mature form is blocked

R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.

FEBS Lett. 225, 188-194, 1987

A;Title: Complete primary structure of the alpha1(1)-chain of human basement membrane (ty

A;Reference number: S00207; MUID:88083584; PMID:3691802

A;Accession: S00207

A;Molecule type: mRNA

A;Residues: 244-530 <SO13>

A;Cross-references: EMBL:Y00706; NID:929548; PIDN:CAA68698.1; PID:G29549

R;Eble, J.A.; Golbig, R.; Mann, K.; Kuehn, K.

EMBO J. 12, 4795-4802, 1993

A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen

A;Reference number: S39614; MUID:94038963; PMID:8223488

A;Accession: S39614

A;Molecule type: protein

A;Residues: 371-554 <EBL>

R;Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984

A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid
A;Reference number: A02863; MUID:85003629; PMID:6434307

A;Accession: A02863

A;Molecule type: protein

A;Residues: 534-718, 'D', 720-835, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 99

A;Experimental source: placenta

R;Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A;Title: Pepsin fragments of human placental basement-membrane collagens showing inter

A;Reference number: S16908; MUID:82005835; PMID:6792033

A;Accession: A58517

A;Molecule type: protein

A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-

R;MacWhirter, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.

Biochemistry 22, 4940-4948, 1983

A;Title: Isolation and characterization of pepsin-solubilized human basement membrane

A;Reference number: S16910; MUID:84053346; PMID:6416291

A;Accession: S16910

A;Molecule type: protein

A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 94

A;Experimental source: placenta

R;Pillalajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;

J. Biol. Chem. 260, 7681-7687, 1985

A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen

A;Reference number: S01466; MUID:85207819; PMID:2581969

A;Accession: S01466

A;Molecule type: mRNA

A;Residues: 1256-1669 <PIH>

A;Cross-references: EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PID:G180424

R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.

Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985

A;Title: Restricted homology between human alpha-1 type IV and other procollagen chain

A;Reference number: S16879; MUID:85216555; PMID:2582422

A;Accession: S16879

A;Molecule type: mRNA

A;Residues: 1259-1669 <BRI>

A;Cross-references: EMBL:M11315; NID:G180817; PIDN:AAAS2042.1; PID:G180818

R;Oberbauer, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss

Eur. J. Biochem. 147, 217-224, 1985

A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha

A;Reference number: A02864; MUID:85127033; PMID:12578961

A;Accession: S19091

A;Molecule type: protein

A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X'

R;Siebold, B.; Deutzmann, R.; Kuehn, K.

Eur. J. Biochem. 176, 617-624, 1988

A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyte

A;Reference number: S02550; MUID:89005112; PMID:2844531

A;Contents: annotation; disulfide bonds

C;Genetics:

A;Gene: GDB:COL4A1

A;Cross-references: GDB:119791; OMIM:120130

A;Map position: 13q34-13q34

A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 23

/; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 1

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

1(IV) chain. The alpha 1(IV) chain contains two alpha 1(IV) chains and one alpha

T29351
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans
N;Alternate names: collagen alpha 2(IV) chain precursor cib-1
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29351
R;Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid F01G12.
A;Reference number: Z20811
A;Accession: T29351
A;Status: preliminary; translated from CE/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1759 <WUX>
A;Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A;Experimental source: strain Bristol N2; clone F01G12
C;Genetics:
A;Gene: CESP:F01G12.5a
A;Map position: X
A;Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3
C;Superfamily: collagen alpha 1(IV) chain

Query Match 76.2%; Score 115; DB 2; Length 1759;
Best Local Similarity 76.9%; Fred. NO. 1.2e-08;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTTPFLFCNVNDVCFASRNDYS 27
|||:|||||:|||||
Db 1582 QRFSTMPFLFCDFNNVCNYSRNDKS 1607

Search completed: April 5, 2004, 07:05:37
Job time : 6.5569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.39952 Seconds
(without alignments)
413.557 Million cell updates/sec

Title: US-10-032-221B-39

Perfect score: 151

Sequence: 1 KQRTTFMPLFCNVNDVCFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 146 | 96.7 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 2 | 145 | 96.0 | 471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 3 | 130 | 86.1 | 754 | 1 CA54 CANFA | Q28247 canis famill |
| 4 | 130 | 86.1 | 1685 | 1 CA54 HUMAN | P29400 homo sapien |
| 5 | 129 | 85.4 | 1669 | 1 CA14 HUMAN | P02462 homo sapien |
| 6 | 129 | 85.4 | 1669 | 1 CA14 MOUSE | P02463 mus musculus |
| 7 | 115 | 76.2 | 1763 | 1 CA24 ASCSU | P27393 ascaris suu |
| 8 | 111 | 73.5 | 1758 | 1 CA24 CAEBL | P17140 caenorhabdi |
| 9 | 107 | 70.9 | 1758 | 1 CA14 CAEBL | P17139 caenorhabdi |
| 10 | 105 | 69.5 | 1691 | 1 CA64 HUMAN | Q14031 homo sapien |
| 11 | 102 | 67.5 | 1707 | 1 CA24 MOUSE | P08122 mus musculus |
| 12 | 102 | 67.5 | 1712 | 1 CA24 HUMAN | P08572 homo sapien |
| 13 | 88 | 58.3 | 623 | 1 CA44 RABIT | P55787 cryptolagus |
| 14 | 88 | 58.3 | 1690 | 1 CA44 HUMAN | P53420 homo sapien |
| 15 | 88 | 58.3 | 1775 | 1 CA14 DROME | P08120 drosophila |
| 16 | 87 | 57.6 | 453 | 1 CA44 BOVIN | Q29442 bos taurus |
| 17 | 49 | 32.5 | 260 | 1 NPA HUMAN | Q31445 homo sapien |
| 18 | 47.5 | 31.5 | 356 | 1 CMO_PIRFU | Q51741 pyrococcus |
| 19 | 46 | 30.5 | 854 | 1 BAL HUMAN | Q08196 homo sapien |
| 20 | 45.5 | 30.1 | 610 | 1 MUTL BORBU | Q51229 borrelia bu |
| 21 | 45.5 | 30.1 | 1743 | 1 TAGC DICDI | Q23868 dictyosteli |
| 22 | 45 | 29.8 | 333 | 1 AMR1 HUMAN | Q39430 homo sapien |
| 23 | 45 | 29.8 | 344 | 1 AMR1 MOUSE | Q39430 mus musculus |
| 24 | 45 | 29.8 | 963 | 1 CUS2 YEAST | P4180 saccharomyc |
| 25 | 45 | 29.8 | 1095 | 1 C25 SACKL | Q02342 saccharomyc |
| 26 | 44.5 | 29.5 | 504 | 1 AMPX VIBPR | Q01693 vibrio prot |
| 27 | 43.5 | 28.8 | 334 | 1 Y092 RICPR | Q02555 rickettsia |
| 28 | 43.5 | 28.8 | 663 | 1 MM02 CHICK | Q30611 gallus gall |
| 29 | 43 | 28.5 | 256 | 1 YNCJ BACSU | P39608 bacillus su |
| 30 | 43 | 28.5 | 301 | 1 RC66 SCHPO | Q94553 schizosacch |
| 31 | 43 | 28.5 | 385 | 1 CHEB BORBU | Q45047 borrelia bu |
| 32 | 43 | 28.5 | 464 | 1 SYE2 COXBU | Q83b16 coxiella bu |
| 33 | 43 | 28.5 | 599 | 1 GP63 LEICH | P15706 leishmania |

34 42.5 28.1 361 1 ALR CORGL
35 42.5 28.1 407 1 NKIR_HUMAN
36 42.5 28.1 407 1 NKIR_MOUSE
37 42.5 28.1 407 1 NKIR_RAT
38 42 27.8 354 1 GBA2 SCHPO
39 42 27.8 358 1 V242_FOWPV
40 42 27.8 505 1 CPDB_MOUSE
41 42 27.8 794 1 SUV5_ARATH
42 42 27.8 834 1 A1IM_YEAST
43 42 27.8 976 1 HMDH_GIBFU
44 42 27.8 1060 1 NKCL_MANSE
45 42 27.8 1260 1 ALS1_CANAL

Q8tau9 corynebacte
P25i03 homo sapien
P30548 mus musculu
P14600 rattus norv
Q04665 schizosacch
Q91428 fowlpox vir
P24457 mus musculu
O82175 arabidopsis
P03875 saccharomyc
Q12577 gibberella
Q25479 manduca sex
P46590 candida alb

ALIGNMENTS

RESULT 1
CA34_HUMAN
ID CA34_HUMAN STANDARD; PRT; 1670 AA.
AC Q01955; Q9BOT2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
[1]
RN RP
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Readers S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues";
RL J. Biol. Chem. 269:23013-23017(1994).
[2]
RN RP
RP REVISIONS.
RA Leinonen A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN RP
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
PRO-574; GLU-1269 AND PRO-1474
RX MEDLINE=21064696; PubMed=1134255;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
[4]
RN RP
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:15780-15784(1992).
[5]
RN RP
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1882840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Readers S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen";
RL Am. J. Hum. Genet. 49:545-554(1991).
[6]
RN RP
RP SEQUENCE OF 1331-1670 FROM N.A.
RX TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;


```

Query Match          96.7%; Score 145; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVCFASRNDYS 27
    |||||
Db 1495 QRFTTTPFLFCNVNDVCFASRNDYS 1520
    |||||

RESULT 2
CA34 BOVIN
ID CA34 BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Reiders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RL encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type

```

```

IV collagens.
-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M63139; AAA62708.1; -.
DR PIR; A39024; A39024.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 4.
DR ProDom; PD03923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT DOMAIN 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match          96.0%; Score 145; DB 1; Length 471;
Best Local Similarity 96.2%; Pred. No. 6.6e-14;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVCFASRNDYS 27
    |||||
Db 296 QRFTTTPFLFCNVNDVCFASRNDYS 321
    |||||

RESULT 3
CA54 CANFA
ID CA54 CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sanroyed; TISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thorne P.S., Marrano P., Bauman R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-

```


J. Biol. Chem. 264:13565-13571 (1989).
[2]
SEQUENCE OF 46-1257 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=8803584; PubMed=3691802;
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
RX "Complete primary structure of the alpha 1-chain of human basement
RT membrane (type IV) collagen.";
RL FBBS Lett. 225:188-194 (1987).
[3]
SEQUENCE OF 1-943 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=88029471; PubMed=3311751;
RA Brazel D., Oberbaumer I., Dieringer H., Babel W., Glanville R.W.,
RX Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
basement membrane collagen (type IV) reveals 21 non-triplet
RT interruptions located within the collagenous domain.";
RL Eur. J. Biochem. 168:529-536 (1987).
[4]
SEQUENCE OF 28-243.
RP MEDLINE=86004708; PubMed=4043082;
RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
RX "Amino acid sequence of the N-terminal aggregation and cross-linking
RT region (VS domain) of the alpha 1 (IV) chain of human basement
membrane collagen.";
RL Eur. J. Biochem. 152:213-219 (1985).
[5]
SEQUENCE OF 534-1447.
RP MEDLINE=8503629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RX "Structure of human-basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1 (IV) chain.";
RL Eur. J. Biochem. 143:545-556 (1984).
[6]
SEQUENCE OF 1256-1669 FROM N.A.
RP MEDLINE=85207819; PubMed=2581969;
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RX Cheung M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha 1 (IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687 (1985).
[7]
SEQUENCE OF 1259-1669 FROM N.A.
RP MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RX Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653 (1985).
[8]
SEQUENCE OF 1-28 FROM N.A.
RP MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hosikka S.L., Tryggvason K.;
RX "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220 (1988).
[9]
SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RP TISSUE=Placenta;
RC MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RX "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624 (1988).
-!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
-!- PMW: Lysines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
-!- PMW: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PMW: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M26576; AAA53098.1; JOINED.
DR EMBL; J04217; AAA53098.1; JOINED.
DR EMBL; M26550; AAA53098.1; JOINED.
DR EMBL; M26540; AAA53098.1; JOINED.
DR EMBL; M26542; AAA53098.1; JOINED.
DR EMBL; M26543; AAA53098.1; JOINED.
DR EMBL; M26544; AAA53098.1; JOINED.
DR EMBL; M26545; AAA53098.1; JOINED.
DR EMBL; M26546; AAA53098.1; JOINED.
DR EMBL; M26547; AAA53098.1; JOINED.
DR EMBL; M26537; AAA53098.1; JOINED.
DR EMBL; M26538; AAA53098.1; JOINED.
DR EMBL; M26548; AAA53098.1; JOINED.
DR EMBL; M26549; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.
DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26539; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.
DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26536; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26541; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA68698.1; -
DR EMBL; X05561; CAA29075.1; -
DR EMBL; M10940; AAA52006.1; -
DR EMBL; M11315; AAA52042.1; -
DR PIR; S16876; CGHU4B.
Genew; HGNC:2202; COL4A1.

FT isoform II).
FT /FTIDVSP_001159.
SQ SEQUENCE 1763 AA, 168526 MW, 304F528BC06AABDD CRC64;
Query Match 76.2%; Score 115; DB 1; Length 1763;
Best Local Similarity 80.0%; Pred. No. 6.6e-09;
Matches 20; Conservative 1; Mismatches 0; Gaps 0;
QY 3 RFTTMEPLFCVNDVNCNFRNDYS 27
DB 1584 RFSTMEPLFCVNDVNCNFRNDKS 1608
RESULT 8
CA24 CAEEL
ID CA24 CAEEL STANDARD; PRT; 1758 AA.
AC P17140; Q19098; Q19099;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (Lethal protein 2).
GN LRT-2 OR CLB-1 OR F01G12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene";
RL J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RX MEDLINE=9008929; PubMed=2793871;
RA Wu X., Le T.T.;
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANTS.
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities";
RL EMBO J. 13:3278-3285(1994).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC Vital for embryonic development.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I; Synonyms=a;
CC IsoId=P17140-1; Sequence=Displayed;
CC Name=II; Synonyms=b;
CC IsoId=P17140-2; Sequence=VSP_001160;
CC -!- DEVELOPMENTAL STAGE: Isoform I is predominant in embryos and
CC isoform II is predominant in the larvae and adults.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.

-!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z22964; CAA80536.1; -
EMBL; Z22964; CAA80537.1; -
EMBL; J05066; AAR27989.1; -
EMBL; U22327; AAG64312.1; ALT_SEQ.
EMBL; U53342; AAA96215.1; -
EMBL; U53342; AAA96216.1; -
PIR; T29350; T29350.
DR WormPep; F01G12.5a; CE04334.
DR WormPep; F01G12.5b; CE04335.
DR GO; GO:0005587; C:collagen type IV; IMP.
DR GO; GO:0030020; F:extracellular matrix structural constituent. .; IMP.
DR GO; GO:0016043; P:cell organization and biogenesis; NAS.
DR InterPro; IPR008161; Clg_Helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 6.
DR SMART; SM00111; C4; 2.
KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
KW Alternative splicing; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1758
FT DOMAIN 27 42
FT DOMAIN 42 1527
FT DOMAIN 1528 1758
FT DISULFID 1546 1635
FT DISULFID 1579 1632
FT DISULFID 1591 1597
FT DISULFID 1654 1750
FT DISULFID 1688 1747
FT DISULFID 1700 1707
FT CARBOHYD 248 248
FT VARSPLIC 229 264
FT VARIANT 48 48
FT VARIANT 366 366
FT VARIANT 570 570
FT VARIANT 588 588
FT VARIANT 597 597
FT VARIANT 690 690
FT VARIANT 690 690
FT VARIANT 737 737
FT VARIANT 877 877
FT VARIANT 904 904
FT VARIANT 1003 1003
FT VARIANT 1125 1125
FT VARIANT 1152 1152
FT VARIANT 1286 1286
FT VARIANT 1604 1604
FT CONFLICT 1682 1682
SQ SEQUENCE 1758 AA; 167750 MW; 97EE3F3DBB2D2AC5 CRC64;
Query Match 73.5%; Score 111; DB 1; Length 1758;
Best Local Similarity 73.1%; Pred. No. 2.5e-08;

OC Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eye, and Kidney;
RX MEDLINE=94171779; PubMed=9125972;
RA Ohashi T., Sugimoto M., Mattei M.-G., Nimomiya Y.,
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RL J. Biol. Chem. 269:7520-7526(1994).
[2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=9175748;
RA Zhou J., Ding M., Zhao Z., Readers S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199(1994).
[3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110
RX MEDLINE=36299642; PubMed=8661006;
RA Zhang X., Zhou J., Readers S.T., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479(1996).
[4]
RP SEQUENCE FROM N.A.
RA Bird C., Grahm D., Lawlor S., Wilson S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=9356449;
RA Zhou J., Mochizuki T., Smets H., Antignac C., Laurila P.,
de Paape A., Tryggvason K., Readers S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169(1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC

CC EMBL; D21337; BAA04809.1; -
DR EMBL; U04845; AAA19569.2; -
DR EMBL; U47004; AAB19038.1; -
DR EMBL; U46959; AAB19038.1; JOINED.
DR EMBL; U46961; AAB19038.1; JOINED.
DR EMBL; U46962; AAB19038.1; JOINED.
DR EMBL; U46963; AAB19038.1; JOINED.
DR EMBL; U46964; AAB19038.1; JOINED.
DR EMBL; U46965; AAB19038.1; JOINED.
DR EMBL; U46966; AAB19038.1; JOINED.
DR EMBL; U46967; AAB19038.1; JOINED.
DR EMBL; U46968; AAB19038.1; JOINED.
DR EMBL; U46969; AAB19038.1; JOINED.
DR EMBL; U46970; AAB19038.1; JOINED.
DR EMBL; U46971; AAB19038.1; JOINED.
DR EMBL; U46972; AAB19038.1; JOINED.
DR EMBL; U46973; AAB19038.1; JOINED.
DR EMBL; U46974; AAB19038.1; JOINED.
DR EMBL; U46975; AAB19038.1; JOINED.
DR EMBL; U46976; AAB19038.1; JOINED.
DR EMBL; U46977; AAB19038.1; JOINED.
DR EMBL; U46978; AAB19038.1; JOINED.
DR EMBL; U46979; AAB19038.1; JOINED.
DR EMBL; U46980; AAB19038.1; JOINED.
DR EMBL; U46981; AAB19038.1; JOINED.
DR EMBL; U46982; AAB19038.1; JOINED.
DR EMBL; U46983; AAB19038.1; JOINED.
DR EMBL; U46984; AAB19038.1; JOINED.
DR EMBL; U46985; AAB19038.1; JOINED.
DR EMBL; U46986; AAB19038.1; JOINED.
DR EMBL; U46987; AAB19038.1; JOINED.
DR EMBL; U46988; AAB19038.1; JOINED.
DR EMBL; U46989; AAB19038.1; JOINED.
DR EMBL; U46990; AAB19038.1; JOINED.
DR EMBL; U46991; AAB19038.1; JOINED.
DR EMBL; U46992; AAB19038.1; JOINED.
DR EMBL; U46993; AAB19038.1; JOINED.
DR EMBL; U46994; AAB19038.1; JOINED.
DR EMBL; U46995; AAB19038.1; JOINED.
DR EMBL; U46996; AAB19038.1; JOINED.
DR EMBL; U46997; AAB19038.1; JOINED.
DR EMBL; U46998; AAB19038.1; JOINED.
DR EMBL; U46999; AAB19038.1; JOINED.
DR EMBL; U47000; AAB19038.1; JOINED.
DR EMBL; U47001; AAB19038.1; JOINED.
DR EMBL; U47002; AAB19038.1; JOINED.
DR EMBL; U47003; AAB19038.1; JOINED.
DR EMBL; U47004; AAB19038.1; JOINED.
DR EMBL; U46960; AAB19039.1; JOINED.
DR EMBL; U46961; AAB19039.1; JOINED.
DR EMBL; U46962; AAB19039.1; JOINED.
DR EMBL; U46963; AAB19039.1; JOINED.
DR EMBL; U46964; AAB19039.1; JOINED.
DR EMBL; U46965; AAB19039.1; JOINED.
DR EMBL; U46966; AAB19039.1; JOINED.
DR EMBL; U46967; AAB19039.1; JOINED.
DR EMBL; U46968; AAB19039.1; JOINED.
DR EMBL; U46969; AAB19039.1; JOINED.
DR EMBL; U46970; AAB19039.1; JOINED.
DR EMBL; U46971; AAB19039.1; JOINED.
DR EMBL; U46972; AAB19039.1; JOINED.
DR EMBL; U46973; AAB19039.1; JOINED.
DR EMBL; U46974; AAB19039.1; JOINED.
DR EMBL; U46975; AAB19039.1; JOINED.
DR EMBL; U46976; AAB19039.1; JOINED.
DR EMBL; U46977; AAB19039.1; JOINED.
DR EMBL; U46978; AAB19039.1; JOINED.
DR EMBL; U46979; AAB19039.1; JOINED.
DR EMBL; U46980; AAB19039.1; JOINED.
DR EMBL; U46981; AAB19039.1; JOINED.
DR EMBL; U46982; AAB19039.1; JOINED.
DR EMBL; U46983; AAB19039.1; JOINED.

```

DR ENBL; U46984; AAB19039.1; JOINED.
DR DR ENBL; U46985; AAB19039.1; JOINED.
DR DR ENBL; U46986; AAB19039.1; JOINED.
DR DR ENBL; U46987; AAB19039.1; JOINED.
DR DR ENBL; U46988; AAB19039.1; JOINED.
DR DR ENBL; U46989; AAB19039.1; JOINED.
DR DR ENBL; U46990; AAB19039.1; JOINED.
DR DR ENBL; U46991; AAB19039.1; JOINED.
DR DR ENBL; U46992; AAB19039.1; JOINED.
DR DR ENBL; U46993; AAB19039.1; JOINED.
DR DR ENBL; U46994; AAB19039.1; JOINED.
DR DR ENBL; U46995; AAB19039.1; JOINED.
DR DR ENBL; U46996; AAB19039.1; JOINED.
DR DR ENBL; U46997; AAB19039.1; JOINED.
DR DR ENBL; U46998; AAB19039.1; JOINED.
DR DR ENBL; U46999; AAB19039.1; JOINED.
DR DR ENBL; U47000; AAB19039.1; JOINED.
DR DR ENBL; U47001; AAB19039.1; JOINED.
DR DR ENBL; U47002; AAB19039.1; JOINED.
DR DR ENBL; U47003; AAB19039.1; JOINED.
DR DR ENBL; AL034369; CA22265.1; -.
DR DR ENBL; AL109943; CAB89263.1; -.
DR DR ENBL; AL136080; CAB96748.1; -.
DR DR ENBL; AL031177; CA220120.1; -.
DR DR ENBL; L22763; AAA16338.1; -.
DR DR PIR; A54122; CGH06B.
DR DR Genew; HGNC:2208; COL4A6.
DR DR NM; 303631; -.
DR DR GO; GO:0005587; C:collagen type IV; NAS.
DR DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
DR DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
DR DR InterPro; IPR008161; Clg helix.
DR DR InterPro; IPR008160; Collagen.
DR DR InterPro; IPR001442; Procollagn4_C.
DR DR Pfam; PF01413; C4; 2.
DR DR Pfam; PF01391; Collagen; 23.
DR DR ProDom; PD000007; Clg helix; 4.
DR DR ProDom; PD003923; ProcollagnC4; 1.
DR DR SMART; SM00111; C4; 2.
DR KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.

Query Match 69.58; Score 105; DB 1; Length 1691;
Best Local Similarity 64.08; Pred. No. 1.8e-07;
Matches 16; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy 3 RFTTMTPLFCNVNDVNCVFASRNDS 27
Db 1518 RFSTMTPIYCNINEVCHYARNDS 1542

RESULT 11
CA24 MOUSE ID CA24 MOUSE STANDARD; PRT; 1707 AA.
AC P08122; G61375;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; PubMed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
RA Pihlajaniemi T., Kurkinen M.;

```

```

RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen.";
RL J. Biol. Chem. 264:6318-6324(1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
RA Deutzmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RT chain and its comparison with the alpha 1(IV) chain.";
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
RT "cDNA and protein sequence of the NCI domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1(IV).";
RL FEBS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen.";
RL FEBS Lett. 206:29-32(1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839906;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RT alpha 2(IV) collagen gene.";
RL Nature 317:177-179(1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----

```


RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RL of basement-membrane type IV collagen. ;
CC Eur. J. Biochem. 176:617-624 (1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire',
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCL) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCL domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or
CC send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05562; CAA29076.1; -
DR EMBL; X05810; CAA29098.1; -
DR EMBL; J02760; AAA58422.1; -
DR EMBL; M36963; AAA53099.1; -
DR EMBL; X12784; CAA31275.1; -
DR EMBL; J04217; AAA53097.1; -
DR PIR; A32024; CGHU2B.
DR Genew; HGNC:2203; COL4A2.
DR MIM; 120090; -
DR GO; GO:0005587; C:collagen type IV; TAS.
DR GO; GO:005201; F:extracellular matrix structural constituent; TAS.
DR GO; GO:0030198; F:extracellular matrix organization and bioge. ; NAS.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; Clg_helix; 7.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25
FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 184 1494 TRIPLE-HELICAL REGION.
FT DOMAIN 1485 1712 NONHELICAL REGION (NCL).
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
FT DISULFID 1549 1590 BY SIMILARITY.
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
FT DISULFID 1646 1708 OR 1708 (BY SIMILARITY).
FT DISULFID 1658 1665 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
FT CONFLICT 471 471 R -> P (IN REF. 2).
FT CONFLICT 683 683 A -> G (IN REF. 2).
FT CONFLICT 1575 1575 M -> I (IN REF. 5).
FT CONFLICT 1663 1663 G -> H (IN REF. 9).
FT CONFLICT 1701 1701 H -> G (IN REF. 9).
SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match 67.5%; Score 102; DB 1; Length 1712;
Best Local Similarity 72.0%; Pred. No. 5.2e-07;

Matches 18; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 3 RETTPEFLFCNVNDVCNEASNDYS 27
DB 1540 RFTSTPEFLFCNVNDVCNYASRNDKS 1564
RESULT 13
CA44_RABIT
ID CA44_RABIT STANDARD; PRT; 623 AA.
AC P55787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Watted M.-G., Ninomiya Y.;
RT 'Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.';
RL J. Biol. Chem. 267:23753-23758 (1992).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire',
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCL) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCL domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or
CC send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01477; -; NOT_ANNOTATED_CDS.
DR PIR; A45137; A45137.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON TER 1 1
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT DOMAIN 393 623 NONHELICAL REGION (NCL).
FT DISULFID 413 502 OR 499 (BY SIMILARITY).
FT DISULFID 446 499 OR 502 (BY SIMILARITY).


```
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR001442; Procollagen4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 21.
DR ProDom: PD000007; Clg_helix; 3.
DR ProDom: PD003923; ProcollagenC4; 1.
DR SMART: SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1690 COLLAGEN ALPHA 4 (IV) CHAIN.
FT DOMAIN 35 64 7S DOMAIN.
FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
FT DOMAIN 1460 1690 NONHELICAL REGION (NC1).
FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1206 1207 CLEAVAGE (BY COLLAGENASE)
BY SIMILARITY.
FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
FT DISULFID 1525 1531 BY SIMILARITY.
FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
FT DISULFID 1634 1641 BY SIMILARITY.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 MISSING (in AS).
FT VARIANT 441 446 /FTID=VAR_008149.
FT VARIANT 545 545 G -> A (in GDSNP:1800516).
FT VARIANT 570 570 /FTID=VAR_008149.
FT VARIANT 570 570 E -> Q.
FT VARIANT 897 897 /FTID=VAR_008150.
FT VARIANT 897 897 G -> E (in FBH).
FT VARIANT 931 931 /FTID=VAR_001912.
FT VARIANT 931 931 A -> T.
FT VARIANT 1004 1004 /FTID=VAR_008151.
FT VARIANT 1004 1004 L -> P (in dBSNP:1800517).
FT VARIANT 1030 1030 /FTID=VAR_008152.
FT VARIANT 1030 1030 G -> V (in AS).
FT VARIANT 1201 1201 /FTID=VAR_008153.
FT VARIANT 1201 1201 G -> S (in AS).
FT VARIANT 1402 1402 /FTID=VAR_001913.
FT VARIANT 1402 1402 P -> S.
FT VARIANT 1572 1572 /FTID=VAR_008154.
FT VARIANT 1572 1572 P -> L (in AS).
FT CONFLICT 1659 1660 /FTID=VAR_008155.
FT CONFLICT 1659 1660 LQ -> FE (IN REF. 3).
SQ SEQUENCE 1690 AA; 164095 MW; E1E72F293A72BARE CRC64;
Query Match 56.3%; Score 88; DB 1; Length 1690;
Best Local Similarity 54.2%; Pred. No. 5.8e-05;
Matches 13; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
QY 4 FTTPMPLFCNNVDVCFNFSNDYS 27
Db 1517 FSTLPAYCNHQVCHYQNRDS 1540
RESULT 15
CA14_DROME STANDARD; PRT; 1775 AA.
AC P08130;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
```

```
DE GN Collagen alpha 1(IV) chain precursor.
OS CG25C OR DCG1.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Eohydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89054012; PubMed=3142875;
RA Blumberg B., Mackrell A.J., Fessler J.H.;
RT "Drosophila basement membrane procollagen alpha 1(IV). II. Complete
RT cDNA sequence, genomic structure, and general implications for
RT supramolecular assemblies.";
RL J. Biol. Chem. 263:18328-18337(1988).
RN [2]
RN SEQUENCE FROM N.A.
RX Blumberg B.;
RA Thesis (1987), University of California / Los Angeles, U.S.A.
RN [3]
RN SEQUENCE FROM N.A.
RX Mackrell A.J.;
RA Thesis (1992), University of California / Los Angeles, U.S.A.
RN [4]
RN SEQUENCE OF 1065-1775 FROM N.A.
RX MEDLINE=87194801; PubMed=3106346;
RA Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,
RA Natzele J.E., Fessler J.H.;
RT "Basement membrane procollagen IV and its specialized carboxyl domain
RT are conserved in Drosophila, mouse, and human.";
RL J. Biol. Chem. 262:5947-5950(1987).
RN [5]
RN SEQUENCE OF 1355-1775 FROM N.A.
RX MEDLINE=87246644; PubMed=3109906;
RA Cecchini J.P., Knibiehler B., Mirre C., Le Parco Y.;
RT "Evidence for a type-IV-related collagen in Drosophila melanogaster.
RT Evolutionary constancy of the carboxyl-terminal noncollagenous
RT domain.";
RL Eur. J. Biochem. 165:587-593(1987).
RN [6]
RN SEQUENCE OF 762-1230 FROM N.A.
RX MEDLINE=82197577; PubMed=6210912;
RA Monson J.M., Natzele J., Friedman J., McCarthy B.J.;
RT "Expression and novel structure of a collagen gene in Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M23704; AAA28404.1; -.
CC EMBL; M96575; AAB59184.1; -.
CC EMBL; J02727; AAA28423.1; -.
DR
DR
```

DR EMBL; M28334; AAA28422.1; -.
DR EMBL; V00200; CAA23486.2; -.
DR PIR; A31893; A31893
DR FlyBase; FBgn000299; Cg25C.
DR GO; GO:000587; C:collagen type IV; NAS.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PDOC00007; C1g_helix; 9.
DR ProDom; PDOC03923; ProcollagnC4; 1.
DR SMART; SMO0111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 ?
FT CHAIN ? 1775 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT DOMAIN ? 1544 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 1545 1775 TRIPLE-HELICAL REGION.
FT DISULFID 1569 1655 NONHELICAL REGION (NC1).
FT DISULFID 1599 1652 OR 1652 (BY SIMILARITY).
FT DISULFID 1611 1617 OR 1655 (BY SIMILARITY).
FT DISULFID 1674 1770 BY SIMILARITY.
FT DISULFID 1708 1767 OR 1767 (BY SIMILARITY).
FT DISULFID 1720 1727 OR 1770 (BY SIMILARITY).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 948 948 L -> S (IN REF. 6).
FT CONFLICT 997 997 S -> T (IN REF. 6).
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).
FT CONFLICT 1373 1373 T -> I (IN REF. 5).
FT CONFLICT 1496 1496 L -> R (IN REF. 5).
FT CONFLICT 1507 1511 ETGNV -> RAGOR (IN REF. 5).
FT CONFLICT 1529 1529 E -> K (IN REF. 5).
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DESAB23149525CD CRC64;

Query Match 58.3%; Score 88; DB 1; Length 1775;
Best Local Similarity 65.2%; Pred. No. 6.1e-05;
Matches 15; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 RPTMPFLFCNVNDYCNFSRND 25
Db 1602 RFSTLPVLSCGQNVNVCYASRND 1624

Search completed: April 5, 2004, 06:59:39
Job time : 4.39952 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 16.2131 Seconds
(without alignments)
525.440 Million cell updates/sec

Title: US-10-032-221b-39

Perfect score: 151

Sequence: 1 KQRTTTPFLFCVNDVCFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 146 | 96.7 | 212 | Q28512 | Q28512 macaca mula |
| 2 | 146 | 96.7 | 245 | Q9NYC4 | Q9NYC4 homo sapien |
| 3 | 145 | 96.0 | 203 | Q29032 | Q29032 sus scrofa |
| 4 | 145 | 96.0 | 203 | Q28682 | Q28682 oryctolagus |
| 5 | 145 | 96.0 | 212 | Q28567 | Q28567 ovis aries |
| 6 | 140 | 92.7 | 161 | Q61430 | Q61430 mus musculus |
| 7 | 140 | 92.7 | 210 | Q28273 | Q28273 canis famil |
| 8 | 140 | 92.7 | 246 | Q61435 | Q61435 mus musculus |
| 9 | 140 | 92.7 | 1669 | 11 Q9QZS0 | Q9QZS0 mus musculus |
| 10 | 136 | 90.1 | 230 | 11 Q63122 | Q63122 rattus norv |
| 11 | 130 | 86.1 | 179 | 11 P70165 | P70165 mus musculus |
| 12 | 130 | 86.1 | 253 | 11 Q61436 | Q61436 mus musculus |
| 13 | 130 | 86.1 | 585 | 11 Q80V57 | Q80V57 mus musculus |
| 14 | 130 | 86.1 | 799 | 11 Q8BNS7 | Q8BNS7 mus musculus |
| 15 | 130 | 86.1 | 886 | 4 Q9NUB7 | Q9NUB7 homo sapien |
| 16 | 130 | 86.1 | 1684 | 6 Q8HYC1 | Q8HYC1 canis famil |

| | | | | | |
|----|-----|------|------|-----------|--------------------|
| 17 | 130 | 86.1 | 1688 | 6 Q856Z2 | Q856Z2 canis famil |
| 18 | 130 | 86.1 | 1891 | 11 Q8ESQ2 | Q8ESQ2 mus musculu |
| 19 | 129 | 85.4 | 225 | 6 Q28271 | Q28271 canis famil |
| 20 | 129 | 85.4 | 226 | 11 Q99LQ8 | Q99LQ8 mus musculu |
| 21 | 129 | 85.4 | 229 | 4 Q8NF88 | Q8NF88 homo sapien |
| 22 | 129 | 85.4 | 229 | 4 Q9NYC5 | Q9NYC5 homo sapien |
| 23 | 129 | 85.4 | 979 | 13 Q919K3 | Q919K3 gallus gall |
| 24 | 129 | 85.4 | 1075 | 4 Q86X41 | Q86X41 homo sapien |
| 25 | 129 | 85.4 | 1821 | 4 Q9H4R9 | Q9H4R9 homo sapien |
| 26 | 120 | 79.5 | 1747 | 5 Q28640 | Q28640 strongyloce |
| 27 | 120 | 79.5 | 1752 | 5 Q07265 | Q07265 strongyloce |
| 28 | 108 | 71.5 | 1802 | 5 Q17163 | Q17163 brugia mala |
| 29 | 105 | 69.5 | 205 | 6 Q28274 | Q28274 canis famil |
| 30 | 105 | 69.5 | 546 | 11 Q99K97 | Q99K97 mus musculu |
| 31 | 105 | 69.5 | 1600 | 4 Q9UEH6 | Q9UEH6 homo sapien |
| 32 | 103 | 69.5 | 1891 | 11 Q8ESQ1 | Q8ESQ1 mus musculu |
| 33 | 102 | 67.5 | 202 | 6 Q28272 | Q28272 canis famil |
| 34 | 102 | 67.5 | 358 | 11 Q91VI3 | Q91VI3 mus musculu |
| 35 | 102 | 67.5 | 673 | 4 Q14052 | Q14052 homo sapien |
| 36 | 99 | 65.6 | 1723 | 5 Q9Q8B1 | Q9Q8B1 hydra atten |
| 37 | 89 | 58.9 | 1761 | 5 Q18407 | Q18407 drosophila |
| 38 | 89 | 58.9 | 1940 | 5 Q9VMV5 | Q9VMV5 drosophila |
| 39 | 88 | 58.3 | 312 | 11 Q64457 | Q64457 mus musculu |
| 40 | 88 | 58.3 | 1682 | 11 Q9QZR9 | Q9QZR9 mus musculu |
| 41 | 88 | 58.3 | 1779 | 5 Q9VMV4 | Q9VMV4 drosophila |
| 42 | 87 | 57.6 | 208 | 6 Q29468 | Q29468 canis famil |
| 43 | 87 | 57.6 | 1024 | 5 Q8T7S4 | Q8T7S4 anopheles g |
| 44 | 84 | 55.6 | 713 | 5 Q9GV24 | Q9GV24 sarcophaga |
| 45 | 52 | 34.4 | 177 | 2 Q9A1Z4 | Q9A1Z4 carsonella |

ALIGNMENTS

RESULT 1

Q28512 PRELIMINARY; PRT; 212 AA.
AC Q28512;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
mammals";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47280; AAA91861.1; -;
DR GO; GO:0005581; C:Collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollag_n4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR PROSITE; PS003923; Procollag_n4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23469 MW; 4BCS74A64E357B64 CRC64;

Query Match 96.7%; Score 146; DB 6; Length 212;
Best Local Similarity 100.0%; Pred. No. 3.2e-15;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
|||||
Db 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 2

Q9NYC4 PRELIMINARY; PRT; 245 AA.
ID Q9NYC4
AC Q9NYC4
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Tumorstatin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,
RA Bricksen M.D., Hopper H., Xiao Y., Stillman I.E., Kalluri R.,
RT "Distinct anti-tumor properties of a type IV collagen domain derived
from basement membrane."
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL: AF258351; AAF72632.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4.C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; ProcollagnC4; 1.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 245
SQ SEQUENCE 245 AA; 26952 MW; 1BES028354D9A57D CRC64;

Query Match 96.7%; Score 146; DB 4; Length 245;

Best Local Similarity 100.0%; Pred. No. 3.7e-15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
|||||
Db 70 QRTTTPFLFCNVNDVCFASRNDYS 95

RESULT 3

Q29032 PRELIMINARY; PRT; 203 AA.
ID Q29032
AC Q29032
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Fusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
mammals."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47284; AAA91882.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4.C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; ProcollagnC4; 1.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 96.0%; Score 145; DB 6; Length 203;

Best Local Similarity 96.2%; Pred. No. 4.5e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
|||||
Db 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 5

Q28567 PRELIMINARY; PRT; 212 AA.
ID Q28567
AC Q28567
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).

```

GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCB1_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47282; AAA91904.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; ProcollagN4_C.
DR Pfam: PF01413; C4; 2.
DR InterPro: IPR000504; RNA_rec_mot.
DR SMART: SMO0111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 212 212
SQ SEQUENCE 212 AA; 23417 MW; 0F58399FCB81BDD8C CRC64;

Query Match 96.0%; Score 145; DB 6; Length 212;
Best Local Similarity 96.2%; Pred. No. 4.7e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCVFASRNDYS 27
|||||:|||||:|||||:|||||:|||||
DB 37 QRTTTPFLFCNVNDVNCVFASRNDYS 62

RESULT 6
Q61430 Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberhaeumer I.;
RT "Cloning of the Ncl domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells."
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X82205; CAA57689.1; -.
DR PIR: S49488; S49488.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; ProcollagN4_C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR SMART: SMO0111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 161 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

```

```

Query Match 92.7%; Score 140; DB 11; Length 161;
Best Local Similarity 92.3%; Pred. No. 2.2e-14;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCVFASRNDYS 27
|||||:|||||:|||||:|||||:|||||
DB 4 QRTTTPFLFCNVNDVNCVFASRNDYS 29

RESULT 7
Q28273 Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCB1_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation."
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL: U50935; AAC48585.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; ProcollagN4_C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; ProcollagN4; 1.
DR SMART: SMO0111; C4; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 210 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA823633D CRC64;

Query Match 92.7%; Score 140; DB 6; Length 210;
Best Local Similarity 92.3%; Pred. No. 2.9e-14;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCVFASRNDYS 27
|||||:|||||:|||||:|||||:|||||
DB 47 QRTTTPFLFCNVNDVNCVFASRNDYS 72

RESULT 8
Q61435 Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal

```

RT laminae: Sequence, distribution, association with laminins, and
RT developmental switches.";
RL J. Cell Biol. 127:879-891(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z35166; CA84529.1; -;
DR PIR; I48302; I48302.
DR MGD; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 92.7%; Score 140; DB 11; Length 246;
Best Local Similarity 92.3%; Pred. No. 3.4e-14;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTMPFLFCNVNDVNCNFSARNDS 27
Db 71 QRFTMPFLFCNVNDVNCNFSARNDS 96

RESULT 9
Q9QZS0
ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.
AC Q9QZS0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 3 collagen IV.
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
Elder F.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169387; AAD50449.1; -;
DR PIR; I48302; I48302.
DR MGD; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 309762B59739A47B2 CRC64;

Query Match 92.7%; Score 140; DB 11; Length 1669;
Best Local Similarity 92.3%; Pred. No. 2.3e-13;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTMPFLFCNVNDVNCNFSARNDS 27
Db 1494 QRFTMPFLFCNVNDVNCNFSARNDS 1519

RESULT 10
Q63122
ID Q63122 PRELIMINARY; PRT; 230 AA.
AC Q63122;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA MEDLINE=98210005; PubMed=950634;
RA Ryan J.J., Katbanna I., Mason P.J., Pusey C.D., Turner A.N.;
RT "Sequence analysis of the 'Goodpasture antigen' of mammals.";
RL Nephrol. Dial. Transplant. 13:602-607(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA Turner N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; I47281; AAB72238.2; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 90.1%; Score 136; DB 11; Length 230;
Best Local Similarity 92.3%; Pred. No. 1.4e-13;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTMPFLFCNVNDVNCNFSARNDS 27
Db 55 QRFTMPFLFCNVNDVNCNFSARNDS 80

RESULT 11
P70165
ID P70165 PRELIMINARY; PRT; 179 AA.
AC P70165;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbauer I.;
RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse
via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and


```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RC MEDLINE=22398257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Stachenko L., Marusina K., Faxner A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski A.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RN "Generation and initial analysis of more than 15,000 full-length human
RN and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN (2)
RN SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RC Strausberg R.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC043317; AAI43317.1; --
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn_4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; procollagnC4; 2.
DR SMART; SM00111; C4; 2.
DR SEQUENCE 585 AA; 59283 MW; 26774FE364F7FD8D CRC64;

Query Match 86.1%; Score 130; DB 11; Length 585;
Best Local Similarity 80.8%; Pred. No. 3.2e-12;
Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVDCNCFASRNDYS 27
DB 411 RRFSTMPFMPFCNNVNCNCFASRNDYS 436
|||||:|||||:|||||:|||||:|||||:
|||||:|||||:|||||:|||||:|||||:

RESULT 14
Q8BNS7 PRELIMINARY; PRT; 799 AA.
ID Q8BNS7
AC Q8BNS7;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Proteolaggen (Fragment).
DE COL4A5.
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RC MEDLINE=22354683; PubMed=12456851;
RX The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RN "Analysis of the mouse transcriptome based on functional annotation of

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 24.3196 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221b-39
Perfect score: 151
Sequence: 1 KQFTTTPFLFCNVNDVCFASNDYS 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 151 | 100.0 | 27 | 6 | ADA20238 |
| 2 | 146 | 96.7 | 79 | 5 | AAU75600 Human typ |
| 3 | 146 | 96.7 | 79 | 6 | ADA20264 Human typ |
| 4 | 146 | 96.7 | 88 | 5 | AAU75608 Human typ |
| 5 | 146 | 96.7 | 88 | 5 | AAU75607 Human typ |
| 6 | 146 | 96.7 | 88 | 6 | ADA20271 Human typ |
| 7 | 146 | 96.7 | 88 | 6 | ADA20272 Human typ |
| 8 | 146 | 96.7 | 124 | 5 | AAU75594 Human typ |
| 9 | 146 | 96.7 | 124 | 6 | ADA20258 Human typ |
| 10 | 146 | 96.7 | 132 | 5 | AAU75597 Human typ |
| 11 | 146 | 96.7 | 132 | 6 | ADA20261 Human typ |
| 12 | 146 | 96.7 | 191 | 5 | AAU75596 Human typ |
| 13 | 146 | 96.7 | 191 | 6 | ADA20260 Human typ |
| 14 | 146 | 96.7 | 211 | 3 | AAU755918 Human typ |
| 15 | 146 | 96.7 | 211 | 5 | ABG79208 Human GP |
| 16 | 146 | 96.7 | 218 | 2 | AAU75164 Human typ |
| 17 | 146 | 96.7 | 218 | 2 | AAU75164 Human typ |
| 18 | 146 | 96.7 | 218 | 3 | AAU756784 Human typ |
| 19 | 146 | 96.7 | 218 | 4 | AAU756784 Human typ |
| 20 | 146 | 96.7 | 232 | 7 | ADC17697 Human typ |
| 21 | 146 | 96.7 | 244 | 5 | ABG79218 Human typ |
| 22 | 146 | 96.7 | 244 | 5 | ABG79219 Human typ |
| 23 | 146 | 96.7 | 244 | 5 | ABG79217 Human typ |
| 24 | 146 | 96.7 | 244 | 5 | AAU75595 Human typ |
| 25 | 146 | 96.7 | 244 | 6 | ADA20225 Human typ |

| | | | | | | |
|----|-----|------|------|---|-----------|-----------|
| 26 | 146 | 96.7 | 245 | 3 | AAU75942 | Human typ |
| 27 | 146 | 96.7 | 245 | 5 | AAU75589 | Human typ |
| 28 | 146 | 96.7 | 254 | 5 | AAU75598 | Human typ |
| 29 | 146 | 96.7 | 268 | 2 | AAU751993 | Type IV C |
| 30 | 146 | 96.7 | 268 | 3 | AAU75555 | Human alp |
| 31 | 146 | 96.7 | 1670 | 7 | ADA47063 | Human pro |
| 32 | 145 | 96.0 | 471 | 2 | AAU79163 | Partial s |
| 33 | 145 | 96.0 | 471 | 2 | AAU44171 | Bovine ty |
| 34 | 145 | 96.0 | 471 | 3 | AAU56783 | Bovine al |
| 35 | 145 | 96.0 | 471 | 4 | AAU56783 | Bovine al |
| 36 | 136 | 90.1 | 230 | 7 | ADA47061 | Rat Prote |
| 37 | 131 | 86.8 | 27 | 6 | ADA20239 | T8-3 pept |
| 38 | 130 | 86.1 | 229 | 7 | ADC17699 | Human typ |
| 39 | 130 | 86.1 | 264 | 2 | AAU75595 | Type IV C |
| 40 | 130 | 86.1 | 264 | 3 | AAU7557 | Human alp |
| 41 | 130 | 86.1 | 309 | 3 | AAU54044 | Human pan |
| 42 | 130 | 86.1 | 772 | 2 | AAU23873 | Human alp |
| 43 | 130 | 86.1 | 772 | 2 | AAU09643 | Human typ |
| 44 | 130 | 86.1 | 1685 | 4 | ABG04839 | Novel hum |
| 45 | 130 | 86.1 | 1693 | 4 | ABG15619 | Novel hum |

ALIGNMENTS

RESULT 1
ADA20238
ID ADA20238 standard; peptide; 27 AA.
XX
AC ADA20238;
XX
DT 20-NOV-2003 (first entry)
XX
DE T8 peptide related to human type IV collagen alpha and angiogenesis.
XX

anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network; C-terminal globular non-collagenous domain; NC1; type IV collagen; cell surface receptor; integrin; angiogenic activity; protein synthesis; cytosolic; gene therapy; T8 peptide; tumstatin; human; type IV collagen alpha 3 chain; mutant; mutein.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"
FT WO2003059257-A2.
XX
PD 24-JUL-2003.
XX
PF 20-DEC-2002; 2002WO-US040938.
XX
PR 21-DEC-2001; 2001US-00032221.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
DR WPI; 2003-597256/55.
XX
PT New peptide, useful for preparing a composition for inhibiting tumor growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Claim 62; Page 45; 240pp; English.
XX
CC This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the T8 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.

XX SQ Sequence 27 AA;

Query Match 100.0%; Score 151; DB 6; Length 27;
 Best Local Similarity 100.0%; Pred. No. 3.5e-16; Indels 0; Gaps 0;
 Matches 27; Conservative 0; Mismatches 0;

QY 1 QRFTTTPFLFCNVNDVCFASRNDYS 27
 |||||
 Db 1 QRFTTTPFLFCNVNDVCFASRNDYS 27

RESULT 2

AAU75600
 ID AAU75600 standard; protein; 79 AA.

XX AC AAU75600;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tum-5.

XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutelin; mutant.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US0000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2002-188037/24.

XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 XX treating disorders involving angiogenesis.

XX PS Example 40; Page; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues
 CC 54-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75589)

XX SQ Sequence 79 AA;

Query Match 96.7%; Score 146; DB 5; Length 79;

Best Local Similarity 100.0%; Pred. No. 6.9e-15;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVCFASRNDYS 27
 |||||
 Db 17 QRFTTTPFLFCNVNDVCFASRNDYS 42

RESULT 3

ADA20264

XX ID ADA20264 standard; protein; 79 AA.

XX AC ADA20264;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-5 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.
 DR N-PSDB; ADA20224.
 XX
 XX
 PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX
 PS Claim 94; SEQ ID NO 26; 240pp; English.
 XX
 CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumor growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC invention is that of tum-5, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.
 XX
 SQ Sequence 79 AA;
 XX
 Query Match 96.7%; Score 146; DB 6; Length 79;
 Best Local Similarity 100.0%; Pred. No. 6.9e-15;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 QRFTMPFLFCNVNDVCFNFSRNDYS 27
 DB 16 QRFTMPFLFCNVNDVCFNFSRNDYS 41
 RESULT 4
 AAU75608
 ID AAU75608 standard; protein; 88 AA.
 XX
 AC AAU75608;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW Tumstatin; angiogenesis; tumour; muten; mutant.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 82
 ET /note= "Wild type Cys substituted with Ala"
 XX
 FN WO200151523-A2.
 PD 19-JUL-2001.
 XX
 XX 08-JAN-2001; 2001WO-US0000565.
 XX
 XX 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 XX
 (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 Kalluri R;
 WPI; 2002-188037/24.
 A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 treating disorders involving angiogenesis.
 Claim 41; Page 153; 205pp; English.
 The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 domain, having one or more of the characteristics selected from: (a) the
 ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 proliferation of endothelial cells; and (c) the ability to cause
 apoptosis of endothelial cells. Also described are the following: (1) use
 of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 analogue or allelic variant in the preparation of a medicament for
 treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 where the angiogenesis is mediated by one or more endothelial cell
 integrins or one or more endothelial cell integrin subunits; or (b) by
 promoting or inducing endothelial cell apoptosis in a tissue, where the
 endothelial cell apoptosis is mediated by one or more endothelial cell
 integrins or one or more endothelial cell integrin subunits; (2) use of
 alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 preparation of a medicament for inhibiting angiogenesis or cell
 proliferation; (3) use of an inhibitor, such as an antibody, antibody
 fragment or peptide of receptor-mediated angiogenesis in the preparation
 of a medicament for treating a proliferative disease in a vertebrate,
 where the disease is characterised by angiogenesis that is mediated by
 receptors to Arresten, Canstatin or Tumstatin and where the receptors
 inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 the presence of a medicament for promoting angiogenesis in a tissue; and
 (5) use of integrins in the preparation of a medicament for promoting or
 inducing angiogenesis or cell proliferation in a tissue. The fragments
 Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 or allelic variants are useful in the preparation of a medicament for
 treating a disorder involving inhibiting angiogenesis in a tissue, where
 the angiogenesis is mediated by one or more endothelial cell integrins or
 one or more endothelial cell integrin subunits; or by promoting or
 inducing endothelial cell apoptosis in a tissue, where the endothelial
 cell apoptosis is mediated by one or more endothelial cell integrins or
 one or more endothelial cell integrin subunits. The medicament is useful
 in inhibiting tumour growth and for the regression of an established
 tumour. The present sequence represents the amino acid sequence of human
 type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
 consists of residues 5-126 of Tumstatin
 Sequence 88 AA;
 Query Match 96.7%; Score 146; DB 5; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.8e-15;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 QRFTMPFLFCNVNDVCFNFSRNDYS 27
 DB 26 QRFTMPFLFCNVNDVCFNFSRNDYS 51
 RESULT 5
 AAU75607
 ID AAU75607 standard; protein; 89 AA.
 XX
 AC AAU75607;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

RESULT 7
ADA20272
ID ADA20272 standard; protein; 88 AA.
XX AC
XX AC
XX DT 20-NOV-2003 (first entry)
XX DE Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX KW metastasis; basement membrane organisation; type IV collagen network;
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human;
XX KW tumstatin 5-125-C-A; mutant; mutein.
XX OS Synthetic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 81
FT /note= "wild-type Cys substituted by Ala at position 125
FT of full-length tumstatin"
XX WO2003059257-A2.
XX PD 24-JUL-2003.
XX PF 20-DEC-2002; 2002WO-US040938.
XX PR 21-DEC-2001; 2001US-00032221.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX WPI; 2003-587256/55.
XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX Claim 94; SEQ ID NO 34; 240pp; English.
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
XX the "tumstatin" protein of the invention which was derived from the amino
XX acid sequence of the alpha 3 chain of human type IV collagen. Note: This
XX sequence (Seq ID33) does not appear in the specification but was created
XX by the indexer from information given in the specification.
XX Sequence 88 AA;
Query Match 96.7%; Score 146; DB 6; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.8e-15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRRTMPFLFCNVNVCNFSRNDYS 27
DB 25 QRRTMPFLFCNVNVCNFSRNDYS 50
RESULT 8
AAU75594
ID AAU75594 standard; protein; 124 AA.
XX AC
XX AC
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX OS Homo sapiens.
XX PN WO200151523-A2.
XX PD 19-JUL-2001.
XX PF 08-JAN-2001; 2001WO-US000565.
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX WPI; 2002-198037/24.
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX Example 33; Page; 205pp; English.
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; or (b) by
XX promoting or inducing endothelial cell apoptosis in a tissue, where the
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in the preparation
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX or allelic variants are useful in the preparation of a medicament for
XX treating a disorder involving inhibiting angiogenesis in a tissue, where

CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting of
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of
 CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in
 CC the specification but is derived from the wild type human Tumstatin
 CC sequence given in figure 18A (see AAU75589)

XX SQ Sequence 124 AA;

Query Match 96.7%; Score 146; DB 5; Length 124;

Best Local Similarity 100.0%; Pred. No. 1.1e-14; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNVCNPFASRNDYS 27
 |||||
 DB 69 QRFTMPFLFCNVNVCNPFASRNDYS 94

RESULT 9

ADA20258

ID ADA20258 standard; protein; 124 AA.

XX AC

XX ADA20258;

XX DT

XX 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor

XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 20; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV

CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"
 CC protein of the invention which was derived from the amino acid sequence
 CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
 CC ID20) does not appear in the specification but was created by the indexer
 CC from information given in the specification.

XX SQ Sequence 124 AA;

Query Match 96.7%; Score 146; DB 6; Length 124;

Best Local Similarity 100.0%; Pred. No. 1.1e-14; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNVCNPFASRNDYS 27
 |||||
 DB 69 QRFTMPFLFCNVNVCNPFASRNDYS 94

RESULT 10

AAU75597

ID AAU75597 standard; protein; 132 AA.

XX AC

XX AAU75597;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tum-2.

XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutant.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2002-188037/24.

XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

XX treating disorders involving angiogenesis.

XX PS Claim 31; Page 152; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,

CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75589)

XX Sequence 132 AA;

Query Match 96.7%; Score 146; DB 5; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
 Db 70 QRFTHMPFLFCNVNDVCFASRNDYS 95

RESULT 11

ID ADA20261
 XX ADA20261 standard; protein; 132 AA.

AC ADA20261;

DT 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

OS Homo sapiens.

XX WO2003059257-A2.

PN 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.

XX Sequence 132 AA;

Query Match 96.7%; Score 146; DB 6; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
 Db 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 12

AAU75596

ID AAU75596 standard; protein; 191 AA.

AC AAU75596;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US0000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
 PT treating disorders involving angiogenesis.

XX Example 32; Page; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2, or
 CC alpha3, alpha4, alpha5, alpha6, beta1 or beta2 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate.
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of
 CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in
 CC the specification but is derived from the wild type human Tumstatin
 CC sequence given in figure 10A (see AAU75589)

XX Sequence 191 AA;

Query Match 96.7%; Score 146; DB 5; Length 191;
 Best Local Similarity 100.0%; Pred. No. 1.8e-14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNDVCFASRNDYS 27
 DB 17 QRFTMPFLFCNVNDVCFASRNDYS 42

RESULT 13

ADA20260
 ID ADA20260 standard; protein; 191 AA.

XX ADA20260;

AC ADA20260;

DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-1 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cyrostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;
 KW tumstatin N53.

XX Homo sapiens.

OS WO2003059257-A2.

PN

XX 24-JUL-2003.
 PD 20-DEC-2002; 2002WO-US040938.
 XX 21-DEC-2001; 2001US-00032221.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA Kalluri R;
 XX WPI; 2003-587256/55.
 DR N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

PS Claim 94; SEQ ID NO 22; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries

CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-

CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular

CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

CC collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cyrostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present

CC sequence is that of tum-1 (tumstatin N53), an abridged form of the

CC "tumstatin" protein of the invention which was derived from the amino

CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This

CC sequence (Seq ID22) does not appear in the specification but was created

CC by the indexer from information given in the specification.

XX Sequence 191 AA;

QY 2 QRFTMPFLFCNVNDVCFASRNDYS 27

DB 16 QRFTMPFLFCNVNDVCFASRNDYS 41

Query Match 96.7%; Score 146; DB 6; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNDVCFASRNDYS 27

DB 16 QRFTMPFLFCNVNDVCFASRNDYS 41

RESULT 14

AAV95918

ID AAV95918 standard; protein; 211 AA.

XX AAV95918;

DT 20-NOV-2000 (first entry)

XX Human Goodpasture antigen DeltaV.

DE Goodpasture antigen; GPdeltaV; goodpasture antigen binding protein; GBPP;

XX human; autoimmune disease; apoptosis; cancer; tumour; therapy.

XX Homo sapiens.

OS WO200050607-A2.

PN 31-AUG-2000.

XX 24-FEB-2000; 2000WO-IB000324.

XX

```
PT 24-FEB-1999; 99US-0121483P.
XX
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2000-572094/53.
DR N-PSDB; AAA50367.
XX
PT Novel Goodpasture antigen binding proteins useful for diagnosing and
XX treating autoimmune disorders, tumor, and preventing cell apoptosis.
PS Claim 36; Page 151-152; 158pp; English.
XX
CC The present sequence is that of human recombinant Goodpasture antigen
CC (GP) Deltav, i.e. an alternative form of human GP resulting from splicing
CC out of exon V. The recombinant protein, lacking the Met-1 residue, was
CC expressed in bacterial pellets using modified vector pET15b carrying
CC GPDeltav cDNA (see AAA50367). The invention relates to novel Goodpasture
CC antigen binding proteins (GPBs, see AA95900-11), which bind to and
CC phosphorylate the unique N-terminal region of human GP, and which are
CC highly expressed in several autoimmune conditions. Claimed methods for
CC treating an autoimmune disorder, cell apoptosis or a tumour involve
CC modifying the expression or activity of GPBP, especially using a GP-
CC derived peptide, such as GPDeltav
XX
SQ Sequence 211 AA;
Query Match 96.7%; Score 146; DB 3; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.1e-14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
DB 69 QRFTHMPFLFCNVNDVCFASRNDYS 94
RESULT 15
ABG79208
ID ABG79208 standard; protein; 211 AA.
XX
AC ABG79208;
XX
DT 15-NOV-2002 (first entry)
XX
DE Human GP protein isoform GPDeltav.
XX
KW Goodpasture antigen binding protein; Goodpasture syndrome;
KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
KW pemphigoid; lichen planus; human.
XX
OS Homo sapiens.
XX
FN WO200261430-A2.
XX
PD 08-AUG-2002.
XX
PF 31-JAN-2002; 2002WO-EP001010.
XX
PR 31-JAN-2001; 2001US-0265249P.
XX
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2002-619280/66.
DR N-PSDB; ABS64491.
XX
PT Identifying candidate compounds for treating autoimmune conditions, e.g.
PT Goodpasture syndrome or lupus, comprises identifying compounds that
```

```
PT reduce phosphorylation of, or formation of conformational isomers of,
PT target proteins.
XX
PS Example 3; Page 199-200; 217pp; English.
XX
CC The invention relates to identifying candidate compounds to treat an
CC autoimmune condition by identifying compounds that reduce phosphorylation
CC of a first target protein (I) which is selected from Goodpasture antigen
CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
CC conformational isomers of the second target protein (II) (selected from
CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
CC NCI domain conformational isomer, which has an amino acid sequence
CC identical to the wild type alpha3 type IV collagen NCI domain, is
CC stabilised by disulphide bonds, and has a molecular weight in a non-
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on
CC chromosome 5q13. The method is useful for treating autoimmune conditions,
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous,
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)
CC domain (also known as the GP antigen) or an MBP isoform
XX
SQ Sequence 211 AA;
Query Match 96.7%; Score 146; DB 5; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.1e-14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
DB 69 QRFTHMPFLFCNVNDVCFASRNDYS 94
Search completed: April 5, 2004, 06:58:32
Job time : 25.3196 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 16.9322 Seconds
(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-39

Perfect score: 151

Sequence: 1 KQRTTTPFLFCNVNDVCFNFSRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 151 | 100.0 | 27 | 14 | US-10-032-221B-39 |
| 2 | 146 | 96.7 | 79 | 14 | US-10-032-221B-26 |
| 3 | 145 | 96.7 | 88 | 14 | US-10-032-221B-33 |
| 4 | 145 | 96.7 | 88 | 14 | US-10-032-221B-34 |
| 5 | 146 | 96.7 | 124 | 14 | US-10-032-221B-20 |
| 6 | 146 | 96.7 | 132 | 14 | US-10-032-221B-23 |
| 7 | 145 | 96.7 | 191 | 14 | US-10-032-221B-22 |
| 8 | 146 | 96.7 | 211 | 14 | US-10-270-877-46 |
| 9 | 146 | 96.7 | 211 | 14 | US-10-270-837-46 |
| 10 | 145 | 96.7 | 232 | 14 | US-10-206-699-304 |
| 11 | 146 | 96.7 | 244 | 14 | US-10-032-221B-10 |
| 12 | 131 | 86.8 | 27 | 14 | US-10-032-221B-40 |
| 13 | 130 | 86.1 | 229 | 9 | US-10-206-699-306 |
| 14 | 130 | 86.1 | 309 | 9 | US-09-925-297-496 |
| 15 | 129 | 85.4 | 229 | 14 | US-10-206-699-302 |

| | | | | | | |
|--------------------|------|----|---------------------|------|-----|----|
| Sequence 2, Appli | 229 | 14 | US-10-032-221B-2 | 85.4 | 129 | 16 |
| Sequence 507, App | 406 | 9 | US-09-925-302-507 | 85.4 | 129 | 17 |
| Sequence 8, Appli | 1659 | 15 | US-10-372-683-8 | 85.4 | 129 | 18 |
| Sequence 42, Appli | 27 | 14 | US-10-032-221B-42 | 84.1 | 127 | 19 |
| Sequence 266, App | 22 | 14 | US-10-206-699-266 | 82.8 | 125 | 20 |
| Sequence 37, Appli | 25 | 14 | US-10-032-221B-37 | 82.8 | 125 | 21 |
| Sequence 38, Appli | 25 | 14 | US-10-032-221B-38 | 77.5 | 117 | 22 |
| Sequence 265, App | 22 | 14 | US-10-206-699-265 | 76.2 | 115 | 23 |
| Sequence 7032, Ap | 1759 | 15 | US-10-369-493-7032 | 76.2 | 115 | 24 |
| Sequence 267, App | 22 | 14 | US-10-206-699-267 | 74.8 | 113 | 25 |
| Sequence 289, App | 22 | 14 | US-10-206-699-289 | 72.8 | 110 | 26 |
| Sequence 29, Appli | 20 | 14 | US-10-032-221B-29 | 72.8 | 110 | 27 |
| Sequence 48095, A | 46 | 9 | US-09-864-761-48095 | 70.9 | 107 | 28 |
| Sequence 5832, Ap | 1744 | 15 | US-10-369-493-5832 | 70.9 | 107 | 29 |
| Sequence 38021, A | 142 | 9 | US-09-864-761-38021 | 69.5 | 105 | 30 |
| Sequence 307, App | 228 | 14 | US-10-206-699-307 | 69.5 | 105 | 31 |
| Sequence 254, App | 18 | 14 | US-10-206-699-254 | 68.9 | 104 | 32 |
| Sequence 260, App | 18 | 14 | US-10-206-699-260 | 68.9 | 104 | 33 |
| Sequence 303, App | 227 | 14 | US-10-206-699-303 | 67.5 | 102 | 34 |
| Sequence 6, Appli | 227 | 14 | US-10-032-221B-6 | 67.5 | 102 | 35 |
| Sequence 518, App | 430 | 9 | US-09-925-302-518 | 67.5 | 102 | 36 |
| Sequence 27, Appli | 459 | 15 | US-10-331-496A-27 | 67.5 | 102 | 37 |
| Sequence 30, Appli | 459 | 15 | US-10-372-683-30 | 67.5 | 102 | 38 |
| Sequence 9, Appli | 1712 | 10 | US-09-961-403-9 | 67.5 | 102 | 39 |
| Sequence 259, App | 18 | 14 | US-10-206-699-259 | 64.9 | 98 | 40 |
| Sequence 270, App | 22 | 14 | US-10-206-699-270 | 64.9 | 98 | 41 |
| Sequence 261, App | 18 | 14 | US-10-206-699-261 | 63.6 | 96 | 42 |
| Sequence 268, App | 22 | 14 | US-10-206-699-268 | 62.9 | 95 | 43 |
| Sequence 253, App | 18 | 14 | US-10-206-699-253 | 62.3 | 94 | 44 |
| Sequence 290, App | 20 | 14 | US-10-206-699-290 | 62.3 | 94 | 45 |

ALIGNMENTS

RESULT 1

US-10-032-221B-39
; Sequence 39, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T8 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length Tumor
; OTHER INFORMATION: atin molecule)
US-10-032-221B-39
Query Match 100.0%; Score 151; DB 14; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.8e-16;

; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match 96.7%; Score 146; DB 14; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
Db 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 6
US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match 96.7%; Score 146; DB 14; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
Db 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 7
US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match 96.7%; Score 146; DB 14; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFCNVNDVCFASRNDYS 27
Db 16 QRFTHMPFLFCNVNDVCFASRNDYS 41

RESULT 8
US-10-270-877-46
; Sequence 46, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match 96.7%; Score 146; DB 14; Length 211;

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 40

; LENGTH: 27

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: T8-3 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8-3)
; OTHER INFORMATION: statin molecule, and serine has been substituted for the cysteine residues at positions 79 and 85)

US-10-032-221B-40

Query Match

Best Local Similarity 86.8%; Score 131; DB 14; Length 27;

Mismatches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

Db 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

RESULT 13

US-10-206-699-306

; Sequence 306, Application US/10206699

; Publication No. US20030100510A1

; GENERAL INFORMATION:

; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.

; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer

; FILE REFERENCE: MBHB 01-1017

; CURRENT APPLICATION NUMBER: US/10/206,699

; CURRENT FILING DATE: 2002-07-26

; PRIOR APPLICATION NUMBER: US 60/308,523

; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: US 60/351,289

; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: US 60/385,362

; PRIOR FILING DATE: 2002-06-03

; NUMBER OF SEQ ID NOS: 307

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 306

; LENGTH: 229

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: alpha 5 chain

US-10-206-699-306

Query Match

Best Local Similarity 86.1%; Score 130; DB 14; Length 229;

Mismatches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

Db 55 RRFSTMPFLFCNINNVCFASRNDYS 80

RESULT 14

US-09-925-297-496

; Sequence 496, Application US/09925297

; Patent No. US20020081659A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: FA105

; CURRENT APPLICATION NUMBER: US/09/925,297

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05989

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 928

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 496

; LENGTH: 309

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (247)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-297-496

Query Match

Best Local Similarity 86.1%; Score 130; DB 9; Length 309;

Mismatches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

Db 135 RRFSTMPFLFCNINNVCFASRNDYS 160

RESULT 15

US-10-206-699-302

; Sequence 302, Application US/10206699

; Publication No. US20030100510A1

; GENERAL INFORMATION:

; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.

; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer

; FILE REFERENCE: MBHB 01-1017

; CURRENT APPLICATION NUMBER: US/10/206,699

; CURRENT FILING DATE: 2002-07-26

; PRIOR APPLICATION NUMBER: US 60/308,523

; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: US 60/351,289

; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: US 60/385,362

; NUMBER OF SEQ ID NOS: 307

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 302

; LENGTH: 229

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: alpha 1 chain

US-10-206-699-302

Query Match

Best Local Similarity 85.4%; Score 129; DB 14; Length 229;

Mismatches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

Db 55 RRFSTMPFLFCNINNVCFASRNDYS 80

Search completed: April 5, 2004, 07:36:06

Job time : 16.9322 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 6.3414 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221B-39
Perfect score: 151
Sequence: 1 QRFTHMPFLFCNVNDVCNFSARNDS 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/1aa/6C_COMB.pep:*
- 6: /cgn2_6/prodata/2/1aa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 146 | 96.7 | 211 | 4 | US-09-512-563C-46 |
| 2 | 146 | 96.7 | 218 | 2 | US-08-399-889-25 |
| 3 | 146 | 96.7 | 218 | 3 | US-09-167-364-25 |
| 4 | 146 | 96.7 | 218 | 3 | US-09-439-897-4 |
| 5 | 146 | 96.7 | 268 | 4 | US-09-589-927-6 |
| 6 | 146 | 96.7 | 268 | 4 | US-09-277-665-6 |
| 7 | 146 | 96.7 | 268 | 4 | US-09-589-987-6 |
| 8 | 145 | 96.0 | 471 | 2 | US-08-399-889-24 |
| 9 | 145 | 96.0 | 471 | 3 | US-09-167-364-24 |
| 10 | 145 | 96.0 | 471 | 3 | US-09-439-897-2 |
| 11 | 130 | 86.1 | 264 | 4 | US-09-589-927-10 |
| 12 | 130 | 86.1 | 264 | 4 | US-09-277-665-10 |
| 13 | 130 | 86.1 | 264 | 4 | US-09-589-987-10 |
| 14 | 129 | 85.4 | 260 | 4 | US-09-589-927-2 |
| 15 | 129 | 85.4 | 260 | 4 | US-09-277-665-2 |
| 16 | 129 | 85.4 | 260 | 4 | US-09-589-987-2 |
| 17 | 105 | 69.5 | 260 | 4 | US-09-589-927-12 |
| 18 | 105 | 69.5 | 260 | 4 | US-09-277-665-12 |
| 19 | 105 | 69.5 | 260 | 4 | US-09-589-987-12 |
| 20 | 102 | 67.5 | 258 | 4 | US-08-399-889-2 |
| 21 | 102 | 67.5 | 258 | 4 | US-09-277-665-4 |
| 22 | 102 | 67.5 | 258 | 4 | US-09-589-987-4 |
| 23 | 88 | 58.3 | 260 | 4 | US-09-589-927-8 |
| 24 | 88 | 58.3 | 260 | 4 | US-09-277-665-8 |
| 25 | 88 | 58.3 | 260 | 4 | US-09-589-987-8 |
| 26 | 68 | 45.0 | 1634 | 1 | US-08-494-168-2 |
| 27 | 51 | 33.8 | 107 | 3 | US-09-102-528-23 |

28 51 33.8 107 3 US-09-102-528-27 Sequence 27, Appl
29 51 33.8 587 3 US-09-102-528-30 Sequence 30, Appl
30 51 33.8 736 3 US-09-102-528-29 Sequence 29, Appl
31 46 30.5 1117 4 US-09-252-991A-23416 Sequence 23416, A
32 45 29.8 410 4 US-09-543-681A-5407 Sequence 5407, Ap
33 44 29.1 49 1 US-07-865-166A-6 Sequence 6, Appl
34 44 29.1 326 4 US-09-134-000C-4813 Sequence 4813, Ap
35 43.5 28.8 663 4 US-09-194-468A-30 Sequence 30, Appl
36 43 28.5 72 4 US-09-543-681A-8206 Sequence 8206, Ap
37 43 28.5 118 3 US-09-413-814-36 Sequence 36, Appl
38 43 28.5 256 4 US-09-194-146-4 Sequence 4, Appl
39 42.5 28.1 340 1 US-08-118-270-49 Sequence 49, Appl
40 42.5 28.1 340 5 PCT-US93-08528-49 Sequence 49, Appl
41 42.5 28.1 372 1 US-07-337-609-20 Sequence 20, Appl
42 42.5 28.1 372 3 US-08-029-170-20 Sequence 20, Appl
43 42.5 28.1 407 1 US-08-117-965-26 Sequence 26, Appl
44 42.5 28.1 407 2 US-08-390-000A-6 Sequence 6, Appl
45 42.5 28.1 407 5 PCT-US92-06532-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 96.7% Score 146; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVCNFSARNDS 27
DB 69 QRFTHMPFLFCNVNDVCNFSARNDS 94

RESULT 2
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT

```
; ORGANISM: Human
US-08-399-889-25

Query Match
Best Local Similarity 96.7%; Score 146; DB 2; Length 218;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTMPFLFCNVNDVCFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match
Best Local Similarity 96.7%; Score 146; DB 3; Length 218;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTMPFLFCNVNDVCFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match
Best Local Similarity 96.7%; Score 146; DB 3; Length 218;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTMPFLFCNVNDVCFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match
Best Local Similarity 96.7%; Score 146; DB 4; Length 268;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTMPFLFCNVNDVCFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match
Best Local Similarity 96.7%; Score 146; DB 4; Length 268;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTMPFLFCNVNDVCFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match
Best Local Similarity 96.7%; Score 146; DB 4; Length 268;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

| Query Match | Score 130; | DB 4; | Length 264; |
|-----------------------|------------|-------|------------------|
| Best Local Similarity | 86.1% | 80.8% | Pred. No. 8e-12; |

CURRENT APPLICATION NUMBER: US/09/277,665

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.5569 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 QRETTMPFLFSNVNDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 126 | 89.4 | 220 | 2 | B49736 |
| 2 | 126 | 89.4 | 1670 | 1 | CGHU3B |
| 3 | 125 | 88.7 | 471 | 2 | A39024 |
| 4 | 120 | 85.1 | 161 | 2 | S49488 |
| 5 | 120 | 85.1 | 246 | 2 | I48302 |
| 6 | 110 | 78.0 | 253 | 2 | I48304 |
| 7 | 110 | 78.0 | 754 | 2 | A55267 |
| 8 | 110 | 78.0 | 1691 | 1 | S22917 |
| 9 | 109 | 77.3 | 258 | 2 | B61228 |
| 10 | 109 | 77.3 | 1659 | 1 | CGHU4B |
| 11 | 109 | 77.3 | 1869 | 1 | CGMS4B |
| 12 | 100 | 70.9 | 1747 | 2 | A54121 |
| 13 | 100 | 70.9 | 1752 | 2 | A45407 |
| 14 | 95 | 67.4 | 1758 | 2 | T29350 |
| 15 | 95 | 67.4 | 1759 | 2 | T29351 |
| 16 | 95 | 67.4 | 1763 | 2 | S16366 |
| 17 | 91 | 64.5 | 261 | 2 | A34476 |
| 18 | 87 | 61.7 | 1744 | 2 | S40991 |
| 19 | 85 | 60.3 | 1691 | 1 | CGHU6B |
| 20 | 82 | 58.2 | 775 | 2 | A61228 |
| 21 | 82 | 58.2 | 1707 | 2 | A33526 |
| 22 | 82 | 58.2 | 1712 | 1 | CGHU2B |
| 23 | 69 | 48.9 | 1761 | 2 | T13990 |
| 24 | 68 | 48.2 | 312 | 2 | I48303 |
| 25 | 68 | 48.2 | 623 | 2 | A45137 |
| 26 | 68 | 48.2 | 1690 | 1 | CGHU1B |
| 27 | 68 | 48.2 | 1775 | 2 | A31893 |
| 28 | 67 | 47.5 | 453 | 2 | S18804 |
| 29 | 52.5 | 37.2 | 610 | 2 | C70126 |

30 48 34.0 332 2 F82140
31 48 34.0 363 2 T37630
32 48 34.0 509 2 AD0648
33 47.5 33.7 334 2 C71718
34 47 33.3 364 2 A97335
35 46.5 33.0 334 2 B97715
36 46.5 33.0 955 2 E84845
37 46 32.6 155 2 B83124
38 46 32.6 457 2 T33494
39 46 32.6 585 2 S15963
40 45.5 32.3 432 2 S51901
41 45.5 32.3 490 2 T24497
42 45 31.9 261 2 B81823
43 45 31.9 452 2 A25346
44 45 31.9 472 2 AB0907
45 45 31.9 623 2 G95180

ALIGNMENTS

RESULT 1

B49736

collagen alpha 3(IV) chain, medium splice form - human (fragment)

N;Contains: collagen alpha 3(IV) chain, splice form GP-V

C;Species: Homo sapiens (man)

C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text_change 17-Mar-2000

C;Accession: B49736; D49736; S69111

R;Peng, L.; Xia, Y.; Wilson, C.B.

J. Biol. Chem. 269, 2342-2348, 1994

A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.

A;Reference number: A49736; MUID:94124597; PMID:8294492

A;Accession: B49736

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 169-220 <PEN1>

A;Accession: D49736

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 32-220 <FEN2>

A;Cross-references: GB:U02519; NID:G409106; PIDN:AAAL8942.1; PID:G409107

A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank

R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wi

Eur. J. Biochem. 229, 754-760, 1995

A;Title: Characterization and expression of multiple alternatively spliced transcripts

of an antigen and one of its alternative forms.

A;Reference number: S69111; MUID:95278230; PMID:7758473

A;Accession: S69111

A;Molecule type: mRNA

A;Residues: 1-45,169-220 <PEN>

C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.

C;Genetics:

A;Gene: GDB:COL4A3

A;Cross-references: GDB:1283351; OMIM:120070

A;Map position: 2q36-2q37

C;Superfamily: collagen alpha 1(IV) chain

F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status prec

F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status

F;2-220/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 89.4%; Score 126; DB 2; Length 220;

Best Local Similarity 92.3%; Pred. No. 1.5e-11;

Matches 2; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 QRETTMPFLFSNVNDFASRNDYS 27

|||||

Db 78 QRETTMPFLFSNVNDFASRNDYS 103

|||||

RESULT 2

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human
 N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
 C:Species: Homo sapiens (man)
 C>Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 22-Jun-1999
 C:Accession: A54763; A43928; A44043; A45971; A39786
 R:Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Readers, S.T.
 J. Biol. Chem. 269, 23013-23017, 1994
 A>Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
 A:Reference number: A54763; MUID:94364994; PMID:8083201
 A:Accession: A54763
 A:Molecule type: mRNA
 A:Residues: 1-1670 <NAR>
 A:CROSS-references: GB:X80031; NID:G577563; PID:G577564
 A:Experimental source: kidney
 R:Turner, N.; Mason, P. J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
 J. Clin. Invest. 89, 592-601, 1992
 A>Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha 3(IV) chain of the human Goodpasture antigen. The alpha 3(IV) chain of the human Goodpasture antigen demonstrates it to be the alpha 3(IV) chain of the human Goodpasture antigen. The alpha 3(IV) chain of the human Goodpasture antigen demonstrates it to be the alpha 3(IV) chain of the human Goodpasture antigen.
 A:Reference number: A43928; MUID:92147878; PMID:1737849
 A:Accession: A43928
 A:Molecule type: mRNA
 A:Residues: 1331-1524, '1', 1526-1670 <TUR>
 A:CROSS-references: GB:X81379
 A:Experimental source: kidney
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 267, 19780-19784, 1992
 A>Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture antigen.
 A:Reference number: A44043; MUID:93015826; PMID:1400291
 A:Accession: A44043
 A:Molecule type: DNA; mRNA
 A:Residues: 1386-1670 <QUI>
 A:CROSS-references: GB:M92993; NID:gl77895; PIDN:AA21610.1; PID:gl77896
 A>Note: Sequence extracted from NCBI backbone (NCBIP:115597)
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 269, 17358, 1994
 A:Reference number: A44738; MUID:94274734; PMID:8006044
 A:Contents: annotation; erratum; correction to intronic sequence in A44043
 R:Bernal, D.; Quinones, S.; Saus, J.
 J. Biol. Chem. 269, 12090-12094, 1993
 A>Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A:Reference number: A45971; MUID:93280184; PMID:8505332
 A:Accession: A45971
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1427-1444 <BER>
 A>Note: Sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly identified as alpha 3(IV) chain of the human Goodpasture antigen.
 R:Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Readers, S.T.
 Am. J. Hum. Genet. 49, 545-554, 1991
 A>Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of the human Goodpasture antigen.
 A:Reference number: A39786; MUID:91353570; PMID:1882840
 A:Accession: A39786
 A:Molecule type: mRNA
 A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A:CROSS-references: GB:S55790; NID:G234418; PIDN:AA19637.1; PID:G234419
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (Gly-X-Y) are subsequently O-glycosylated.
 C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope of the alpha 3(IV) chain of the human Goodpasture antigen.
 C:Genetics:
 A:Gene: GDB:COL4A3
 A:CROSS-references: GDB:128351; OMIM:120070
 A:Map position: 2q36-q37
 A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
 A>Note: The alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with the alpha 3(IV) chain gene on the plus strand and the alpha 4(IV) chain gene on the minus strand.
 C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3(IV) chains and one alpha 4(IV) chain. The alpha 3(IV) chain is thought to form a heterotrimer of two alpha 3(IV) chains and one alpha 4(IV) chain. The alpha 3(IV) chain is thought to form a heterotrimer of two alpha 3(IV) chains and one alpha 4(IV) chain. The alpha 3(IV) chain is thought to form a heterotrimer of two alpha 3(IV) chains and one alpha 4(IV) chain.
 C:Function:
 A:Description: minor structural component of extracellular basement membrane in kidney glomeruli.
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracellular matrix; signal sequence #status predicted <SIG>
 F1:29-Domain: signal sequence #status predicted <SIG>
 F1:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <NAT>
 F1:29-42/DNA: amino-terminal nonhelical, NH1 <NH1>

F1:43-1438/Region: interrupted helical
 F1:791-793/Region: cell attachment (R-G-D) motif
 F1:996-998/Region: cell attachment (R-G-D) motif
 F1:1154-1156/Region: cell attachment (R-G-D) motif
 F1:1306-1308/Region: cell attachment (R-G-D) motif
 F1:1345-1347/Region: cell attachment (R-G-D) motif
 F1:1432-1434/Region: cell attachment (R-G-D) motif
 F1:1439-1670/DNA: carboxyl-terminal nonhelical, NCI <NC1>
 F1:1451-1551/DNA: collagen IV carboxyl-terminal repeat <CT1>
 F1:1561-1665/DNA: collagen IV carboxyl-terminal repeat <CT2>
 F1:31-33, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1367/Disulfide bonds: interchain #status predicted <COL>
 F1:253/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F1:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F1:1505-1511, 1616-1622/Disulfide bonds: #status predicted
 F1:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
 Query Match 89.4%; Score 126; DB 1; Length 1670;
 Best Local Similarity 92.3%; Pred. No. 1.5e-10;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 QRTTMTPELFNVDVNFASRNDYS 27
 DB 1495 QRTTMTPELFNVDVNFASRNDYS 1520
 RESULT 3
 A39024
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 04-Dec-1992 #sequence revision 04-Dec-1992 #text change 13-Aug-1999
 C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
 R:Morrison, K.E.; Germino, G.G.; Readers, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A>Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the Goodpasture antigen.
 A:Reference number: A39024; MUID:91093146; PMID:1985905
 A:Accession: A39024
 A:Molecule type: mRNA
 A:Residues: 1-471 <MOR>
 A:CROSS-references: EMBL:M63139; NID:gl62886; PIDN:AA62708.1; PID:gl62887
 R:Kotowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 267, 7874-7877, 1992
 A>Title: Localization of the Goodpasture epitope to a novel chain of basement membrane protein.
 A:Reference number: S18432; MUID:87222419; PMID:2438283
 A:Accession: S20672
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-244 <BUT>
 R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A>Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen type IV.
 A:Reference number: S17802; MUID:88330844; PMID:3417661
 A:Accession: S17802
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 265, 5466-5469, 1990
 A>Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type IV collagen.
 A:Reference number: A35167; MUID:90202779; PMID:2318822
 A:Accession: A35167
 A:Molecule type: protein
 A:Residues: 236-255 <GUN>
 R:Gunwar, S.; Ballerster, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; No
 J. Biol. Chem. 266, 15318-15324, 1991
 A>Title: Glomerular basement membrane. Identification of dimeric subunits of the noncollagenous domain.
 A:Reference number: A39419; MUID:91332055; PMID:1869555
 A:Accession: C39419
 A:Molecule type: protein
 A:Residues: 236-255 <GUN>
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; extracellular matrix; signal sequence #status predicted <COL>
 F1:238-Domain: collagenous (fragment) #status predicted <COL>
 F1:239-471/DNA: carboxyl-terminal nonhelical, NCI #status predicted <NC1>
 F1:239-353/DNA: repeat NCI #status predicted <NC1>
 F1:354-471/DNA: repeat NCI #status predicted <NC12>

F:232,238/Modified site: hydroxyproline (Pro) #status experimental
F:306,312,417-423/Disulfide bonds: #status predicted

Query Match 88.7%; Score 125; DB 2; Length 471;
Best Local Similarity 88.5%; Pred. No. 5.1e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFSNVNDVSNFASRNDYS 27
DB 296 QRFTMPFLFCNINNVCFASRNDYS 321

RESULT 4

I48304
collagen alpha 3(IV) chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C:Accession: S49488
R:Oberbaeumer, I.
A:Description: Cloning of the NCI domains fo the minor collagen IV chains of mouse via F
ells.

A:Reference number: S49487
A:Accession: S49488
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-161 <OBE>
A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G559916
C:Superfamily: collagen alpha 1(IV) chain

Query Match 85.1%; Score 120; DB 2; Length 161;
Best Local Similarity 84.6%; Pred. No. 8.4e-11;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFSNVNDVSNFASRNDYS 27
DB 4 QRFTMPFLFCNINNVCFASRNDYS 29

RESULT 5

I48302
collagen alpha 3(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
C:Accession: I48302; S47278
R:Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48302
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <RES>
A:Cross-references: EMBL:Z35166; NID:G535197; PID:G535198
C:Superfamily: collagen alpha 1(IV) chain

Query Match 85.1%; Score 120; DB 2; Length 246;
Best Local Similarity 84.6%; Pred. No. 1.4e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFSNVNDVSNFASRNDYS 27
DB 71 QRFTMPFLFCNINNVCFASRNDYS 96

RESULT 6

I48304
collagen alpha 5(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999
C:Accession: I48304; S47280
R:Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq
A:Reference number: A54979; MUID:95050957; PMID:7962065
A:Accession: I48304
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <RES>
A:Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202
C:Superfamily: collagen alpha 1(IV) chain

Query Match 78.0%; Score 110; DB 2; Length 253;
Best Local Similarity 73.1%; Pred. No. 4.6e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFSNVNDVSNFASRNDYS 27
DB 79 RRFSTWPFMFCNINNVCFASRNDYS 104

RESULT 7

A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267
R:Zheng, K.; Thorne, P.S.; Marrano, P.; Baumal, R.; McIntnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-J
en type IV.

A:Reference number: A55267; MUID:94224868; PMID:8171024

A:Accession: A55267
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-754 <ZHE>
A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548
C:Superfamily: collagen alpha 1(IV) chain

Query Match 78.0%; Score 110; DB 2; Length 754;
Best Local Similarity 73.1%; Pred. No. 1.6e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFSNVNDVSNFASRNDYS 27
DB 587 RRFSTWPFMFCNINNVCFASRNDYS 612

RESULT 8

S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N:Alternate names: procollagen alpha 5(IV) chain
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A3
R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992

A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ident
n Alport syndrome patient
A:Reference number: S22917; MUID:92316923; PMID:1352287
A:Accession: S22917
A:Molecule type: mRNA

A:Residues: 1-967 <ZHO>
A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234
R:Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994

A:Title: Structure of the human type IV collagen COL4A5 gene.
A:Reference number: A54365; MUID:94165049; PMID:8120014

A:Accession: A54365
A:Molecule type: DNA
A:Residues: 1-922 <ZH2>

A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAC27816.1; P
R:Zhou, J.; Mochizuki, T.; Smets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggva
Science 261, 1167-1169, 1993
A:Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited s

A,Cross-references: GB:S75903; NID:G913882; PIDN:AAB33374.1.; PID:G913883
A>Note: permatue termination mutation from a patient with Alport syndrome; one other m
R;Leimink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;
Genomics 17, 485-489, 1993
A>Title: Identification of four novel mutations in the COL4A5 gene of patients with Alp
A:Reference number: 154188; MUID:94010948; PMID:8406498
A:Accession: 154188
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1604-1607, 'VHDAYKC' <L>
A,Cross-references: GB:S65767; MID:G425563; PIDN:AAD13967.1.; PID:G4261667
A>Note: frameshift mutation from a patient with Alport syndrome; five other mutations a
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A5; ATS
A,Cross-references: GDB:120596; OMIM:303630
A:Map position: Xq22-Xq22
A:Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1;
A>Note: the alpha 5 (IV) and alpha 6(IV) chain genes are encoded on opposite strands wit
C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-1

C:Function:
A:Description: minor structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT
F:27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F:42-1462/Region: interrupted helical
F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <Ct1>
F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <Ct2>
F:29,32,38,40,124,451,461,484/Disulfide bonds: interchain #status predicted
F:125/Binding site: carbonyl-dratc (Asn) (covalent) #status predicted
F:1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F:1527-1533,1638-1644/Disulfide bonds: #status predicted
F:1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 78.0%; Score 110; DB 1; Length 1691;
Best Local Similarity 73.1%; Pred.No. 4,1e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDS Y27
:::|||||:::|:::|||||
DB 1517 RRFTSTPFFMFCNNVNCFASRNDS 1542
:::|||||:::|:::|||||

RESULT 9
B61228
collagen alpha 1(IV) chain - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
C:Accession: B61228
R;Yamaguchi, N.; Sato, N.; Ko, J.S.; Niimiya, Y.
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991
A>Title: Cloning of alpa1(IV) and alpa2(IV) collagen cDNAs from rabbit corneal endoth
A:Reference number: A61228; MUID:92010685; PMID:1717398
A:Accession: B61228
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-258 <YAM>
C:Superfamily: collagen alpha 1(IV) chain

Query Match 77.3%; Score 109; DB 2; Length 258;
Best Local Similarity 73.1%; Pred.No. 6,7e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDS Y27
:::|||||:::|:::|||||

Db 84 RKFTMPFLFCNNVNCNFSRNDYS 109

RESULT 10

COHU4B

collagen alpha 1(IV) chain precursor - human
N:Alternate names: procollagen alpha 1(IV) chain

C:Species: Homo sapiens (man)

C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999

C:Accession: S16876; A32117; S02738; S00048; S28826; A23115; S00207; S39614; A02863; A58

R:Soinininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 264, 13565-13571, 1989

A:Title: Structural organization of the gene for the alpha-1 chain of human type IV coll

A:Reference number: S16876; MUID:89340433; PMID:2701944

A:Accession: S16876

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1669 <SO11>

A:Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAAS3098.1; PID:G180803

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988

R:Soinininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988

A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are

A:Reference number: A32690; MUID:89034231; PMID:3182844

A:Accession: A32117

A:Molecule type: DNA

A:Residues: 1-28 <SO12>

A:Cross-references: EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PID:G553233

R:Poerschl, E.; Pollner, R.; Kuehn, K.

EMBO J. 7, 2687-2695, 1988

A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c

A:Reference number: S02738; MUID:89030632; PMID:2846280

A:Accession: S02738

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-6, 'L', 8-28 <POE>

A:Cross-references: EMBL:X12784; NID:G30072

R:Bräzel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.

Eur. J. Biochem. 188, 529-536, 1987

A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem

A:Reference number: S00048; MUID:88029471; PMID:3311751

A:Accession: S00048

A:Molecule type: mRNA

A:Residues: 1-318, 'A', 320-944 <BRA1>

A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067

A:Accession: S25826

A:Molecule type: protein

A:Residues: 271-318, 'A', 320-554 <BRA2>

R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.

Eur. J. Biochem. 152, 213-219, 1985

A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S

A:Reference number: A23115; MUID:86004708; PMID:4043082

A:Accession: A23115

A:Molecule type: protein

A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

A:Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549

R:EBle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.

EMBO J. 12, 4795-4802, 1993

A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen

A:Reference number: S39614; MUID:94038963; PMID:8223488

A:Accession: S39614

A:Molecule type: protein

A:Residues: 371-554 <ESL>

R:Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984

A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid s

A:Reference number: A02863; MUID:85003629; PMID:6434307

A:Accession: A02863

A:Molecule type: protein

A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999;

A:Experimental source: placenta

R:Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A:Title: Peptin fragments of human placental basement-membrane collagens showing inter

A:Reference number: S16908; MUID:82005835; PMID:6792033

A:Accession: A58517

A:Molecule type: protein

A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-1

R:Wachsmuth, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.

Biochemistry 22, 4940-4948, 1983

A:Title: Isolation and characterization of pepsin-solubilized human basement membrane

A:Reference number: S16910; MUID:84053346; PMID:6416291

A:Accession: S16910

A:Molecule type: protein

A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948;

A:Experimental source: placenta

R:Pihtajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;

J. Biol. Chem. 260, 7681-7687, 1985

A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen

A:Reference number: S01466; MUID:85207819; PMID:2581969

A:Accession: S01466

A:Molecule type: mRNA

A:Residues: 1256-1669 <PIH>

A:Cross-references: EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PID:G180424

R:Brinker, J.M.; Gudus, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;

Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985

A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains

A:Reference number: S16879; MUID:85216555; PMID:2582422

A:Accession: S16879

A:Molecule type: mRNA

A:Residues: 1259-1669 <BRI>

A:Cross-references: EMBL:M11315; NID:G180817; PIDN:AAAS2042.1; PID:G180818

R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,

Eur. J. Biochem. 147, 217-224, 1985

A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-

A:Reference number: A02864; MUID:85127033; PMID:2578961

A:Accession: S19091

A:Molecule type: protein

A:Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',

R:Siebold, B.; Deutzmann, R.; Kuehn, K.

Eur. J. Biochem. 176, 617-624, 1988

A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyter

A:Reference number: S02550; MUID:89005112; PMID:2844531

A:Contents: annotation; disulfide bonds

C:Genetics:

A:Gene: COL4A1

A:Cross-references: GDB:119791; OMIM:120130

A:Map position: 13q34-13q34

A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231

/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 1020/1; 1066/3; 1109/1; 1136/1; 11

C:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2

oiations among trimer amino-terminal domains (disulfide and desmosine cross-links), di

r-trimer associations in the interrupted helical domain (with disulfide and desmosine c

C:Function:

A:Description: structural component of extracellular basement membrane

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicat

P:1-26/Domain: signal sequence #status predicted <SIG>

P:29-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>

P:163-1440/Domain: amino-terminal nonhelical, 78 <7SD>

P:29-1669/Domain: interrupted helical <COL>

P:414-452/Region: integrin binding #status experimental

P:597-599/Region: cell attachment (R-G-D) motif

P:917-919/Region: cell attachment (R-G-D) motif

P:968-970/Region: cell attachment (R-G-D) motif

P:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

P:1451-1551/Domain: collagen IV carboxyl-terminal repeat <C1I>

Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
Gene 43, 301-304, 1986
A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeoxynucleotide as a probe.
A;Reference number: A25636; PMID:86301886; PMID:3755692
A;Accession: A25636
A;Molecule type: mRNA
A;Residues: 1149-1396; 'S', 1398-1424 <NAT>
A;Cross-references: EMBL:ML0402; NID:G192286; PIDN:AAA37342.1; PID:G192287
A;Note: the authors translated the codon CAG for residue 1374 as Arg
A;Note: Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla
Eur. J. Biol. Chem. 262, 8496-8499, 1987
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
collagen and human alpha-1(IV) collagen.
A;Reference number: A94680; PMID:87250460; PMID:3597383
A;Accession: A29301
A;Molecule type: mRNA
A;Residues: 1441-1669 <KUR>
A;Cross-references: EMBL:W15932; NID:G192282; PIDN:AAA37340.1; PID:G387115
A;Note: Killen, P.D.; Burbelo, P.D.; Martin, G.R.; Yamada, Y.
Eur. J. Biol. Chem. 263, 12310-12314, 1988
A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequen
A;Reference number: S19079; PMID:88315019; PMID:2842328
A;Accession: S19079
A;Molecule type: DNA
A;Residues: 1-28 <K12>
A;Cross-references: EMBL:J003944; NID:G192673; PIDN:AAA37442.1; PID:G466503
A;Note: Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
Eur. J. Biol. Chem. 263, 19274-19277, 1988
A;Title: Head-to-head arrangement of murine type IV collagen genes.
A;Reference number: A92702; PMID:89066738; PMID:3198626
A;Accession: A32003
A;Molecule type: DNA
A;Residues: 1-28 <KAY>
A;Cross-references: EMBL:J04448; NID:G1922666; PIDN:AAA37437.1; PID:G450449
A;Note: Burbelo, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
A;Reference number: A94220; PMID:89071759; PMID:3200851
A;Accession: A31766
A;Molecule type: DNA
A;Residues: 1-28 <BUR>
A;Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668
A;Note: Sakurai, Y.; Sullivan, M.; Yamada, Y.
Eur. J. Biol. Chem. 261, 6654-6657, 1986
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen gene
A;Reference number: S19094; PMID:86196099; PMID:3009468
A;Accession: S19094
A;Molecule type: DNA
A;Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>
A;Cross-references: EMBL:M13027
A;Note: Ruchshuppan, D.; Timpl, R.; Glanville, R.W.
FEBS Lett. 115, 237-300, 1980
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (t
A;Reference number: S16909; PMID:80246483; PMID:6772473
A;Accession: S16909
A;Molecule type: protein
A;Residues: 940-946; 'G', 948-949; 'G', 951-955; 'G', 957; 1213-1228; 'X', 1230-1234; 'P', 1236-12
A;Cross-references: EMBL:G10000; NID:G192282; PIDN:AAA37340.1; PID:G387115
A;Note: Ruchshuppan, D.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 123, 505-512, 1982
A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial ami
A;Reference number: A25991; PMID:82186723; PMID:6804236
A;Accession: A25991
A;Molecule type: protein
A;Residues: 940-946; 'X', 948-949; 'X', 951-955; 'X', 957-964; 'X', 966-991; 'X', 993-1003; 'X', 10
64; 'X', 1063-1085; 'X', 1067-1080; 'X', 1082-1083; 'X', 1085-1106; 'X', 1108-1115; 'DE', 1118-1119
A;Cross-references: EMBL:G10000; NID:G192282; PIDN:AAA37340.1; PID:G387115
A;Note: Ruchshuppan, D.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A;Title: Subunit structure and assembly of the globular domain of basement-membrane col
A;Reference number: S17801; PMID:84132058; PMID:6698021
A;Accession: S17801

A:Molecule type: protein
A:Residues: 1435-1443 <WEB>

C:Genetics: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3

A:Note: the list of introns may be incomplete

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular mat

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <VAR>

F:128-162/Domain: 7S <7SD>

F:163-1440/Domain: collagenous, triple helix <COL>

F:597-599/Region: cell attachment (R-G-D) motif

F:781-783/Region: cell attachment (R-G-D) motif

F:917-919/Region: cell attachment (R-G-D) motif

F:968-970/Region: cell attachment (R-G-D) motif

F:1441-1669/Domain: carboxyl-terminal nonhelical, NCI <NCI>

F:1441-1552/Region: duplication

F:1553-1669/Region: duplication

F:31.36.39.41.434.467.470/Disulfide bonds: interchain #status predicted

F:126/Binding site: carboxylate (Asn) (covalent) #status predicted

F:971.974.977.986.989.1001.1007.1019.1022.1031.1037.1040.1055.1060.1063.1075.1078.1090.1

92.1298.1310.1313.1322.1337.1346.1349.1422.1425.1431.1437.1440/Modified site: hydroxypro

F:1214.1424/Modified site: 4-hydroxyproline (Pro) #status experimental

F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental

F:1503-1511.1616-1622/Disulfide bonds: #status predicted

Query Match 77.3%; Score 109; DB 1; Length 1669;

Best Local Similarity 73.1%; Pred. No. 5.7e-08;

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNVSFASRNDYS 27

DB 1495 RFSTMPFLFCNNVNCVAFSNDYS 1520

RESULT 12

A54121

collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)

N:Alternate names: collagen alpha 2(IV) chain homolog

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999

C:Accession: A54121; S44317

R:Exposito, J.Y.; Suzuki, H.; Geourjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.

J. Biol. Chem. 269, 13167-13171, 1994

A:Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I

A:Reference number: A54121; MUID:94230414; PMID:8175744

A:Molecule type: mRNA

A:Residues: 1-1747 <EXP>

A:Cross-references: EMBL:X76730; NID:G483606; PIDN:CAA54146.1; PID:G483607

C:Genetics: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3

C:Superfamily: collagen alpha 1(IV) chain

Query Match 70.9%; Score 100; DB 2; Length 1747;

Best Local Similarity 73.1%; Pred. No. 1.4e-06;

Matches 19; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNVSFASRNDYS 27

DB 1576 QRFTHMPFLFCNNVNCVAFSNDYS 1601

RESULT 13

A5407

collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999

C:Accession: A5407; A43903; A23940

R:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.

J. Biol. Chem. 268, 5249-5254, 1993

A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and ana

A:Reference number: A5407; MUID:93186842; PMID:844899

A:Accession: A45407

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1752 <EXP>

A:Note: sequence extracted from NCBI backbone (NCBIP:126841)

R:Wessel, G.M.; Etkin, M.; Benson, S.

Dev. Biol. 148, 261-272, 1991

A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously proc

A:Reference number: A43903; MUID:92038439; PMID:1936564

A:Accession: A43903

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 'P', 633-1537, 'G' <VEN>

A:Cross-references: GB:S64572; NID:G238616; PIDN:AAB20270.1; FID:G238617

A:Note: sequence extracted from NCBI backbone (NCBIN:64572, NCBIP:64573)

R:Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.

Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986

A:Title: Structure and developmentally regulated expression of a Strongylocentrotus pur

A:Reference number: A23940; MUID:86205894; PMID:3458186

A:Accession: A23940

A:Molecule type: DNA

A:Residues: 742-812 <VEN>

A:Cross-references: EMBL:M13206

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

F:162-1523/Region: amino-terminal nonhelical, 7S <7SD>

F:1524-1752/Domain: carboxyl-terminal nonhelical, NCI <NCI>

F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>

F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>

F:129/Modified site: allylsine (Lys) #status predicted

Query Match 70.9%; Score 100; DB 2; Length 1752;

Best Local Similarity 69.2%; Pred. No. 1.4e-06;

Matches 18; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNVSFASRNDYS 27

DB 1578 KRFTMPFLFCNNVNCVAFSNDYS 1603

RESULT 14

T29350

hypothetical protein F01G12.5a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T29350

R:Wu, X.; Le, T.T.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F01G12.

A:Reference number: Z20611

A:Accession: T29350

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1758 <WUX>

A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a

A:Experimental source: strain Bristol N2; clone F01G12

C:Genetics: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3

A:Gene: CESP:F01G12.5a

A:Map position: X

A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/

C:Superfamily: collagen alpha 1(IV) chain

Query Match 67.4%; Score 95; DB 2; Length 1758;

Best Local Similarity 69.2%; Pred. No. 8e-06;

Matches 18; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNVSFASRNDYS 27

DB 1581 QRFTHMPFLFCNNVNCVAFSNDYS 1606

RESULT 15

Mon Apr 5 07:53:14 2004

T29351
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans
N/Alternate names: collagen alpha 2(IV) chain precursor clb-1
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T29351
R/Wu, X.; Le, T.T.
Submitted to the EMBL Data Library, April 1996
A/Description: The sequence of C. elegans cosmid F01G12.
A/Reference number: Z20611
A/Accession: T29351
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1759 <WUX>
A/Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A/Experimental source: Strain Bristol N2, clone F01G12
C/Genetics:
A/Gene: CESP:F01G12.5a
A/Map position: X
A/Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3
C/Superfamily: collagen alpha 1(IV) chain

Query Match 67.4%; Score 95; DB 2; Length 1759;
Best Local Similarity 69.2%; Pred. No. se-06; 4; Indels 0; Gaps 0;
Matches 18; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFTHPELFGNVDVSNFASRNDYS 27
|||:|||||:|:|:|
Db 1582 QRFTHPELFGCDFNVCNYSRNDKS 1607

Search completed: April 5, 2004, 07:05:38
Job time : 6.5569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.39952 Seconds
(without alignments)
413.557 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 KQRTTTFPLFSNVNDVSNFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 126 | 89.4 | 1670 | 1 CA34_HUMAN | Q01955 homo sapien |
| 2 | 125 | 88.7 | 471 | 1 CA34_BOVIN | Q28084 bos taurus |
| 3 | 110 | 78.0 | 754 | 1 CA54_CANFA | Q28247 canis famil |
| 4 | 110 | 78.0 | 1685 | 1 CA54_HUMAN | P29460 homo sapien |
| 5 | 109 | 77.3 | 1669 | 1 CA14_HUMAN | P02462 homo sapien |
| 6 | 109 | 77.3 | 1669 | 1 CA14_MOUSE | P02463 mus musculus |
| 7 | 95 | 67.4 | 1763 | 1 CA24_ASCSU | P27393 ascaris suu |
| 8 | 91 | 64.5 | 1758 | 1 CA24_CAEEL | P17139 caenorhabdi |
| 9 | 87 | 61.7 | 1758 | 1 CA14_CAEEL | P17139 caenorhabdi |
| 10 | 85 | 60.3 | 1691 | 1 CA64_HUMAN | Q14031 homo sapien |
| 11 | 82 | 58.2 | 1707 | 1 CA24_MOUSE | P08122 mus musculus |
| 12 | 82 | 58.2 | 1712 | 1 CA24_HUMAN | P08572 homo sapien |
| 13 | 68 | 48.2 | 623 | 1 CA44_RABIT | P55787 oryctolagus |
| 14 | 68 | 48.2 | 1690 | 1 CA14_HUMAN | P53420 homo sapien |
| 15 | 68 | 48.2 | 1775 | 1 CA14_DROME | P08120 drosophila |
| 16 | 67 | 47.5 | 453 | 1 CA44_BOVIN | Q29442 bos taurus |
| 17 | 52.5 | 37.2 | 610 | 1 MUTL_BORBU | O51229 borrelia bu |
| 18 | 47.5 | 33.7 | 334 | 1 Y032_RICPR | Q92655 rickettsia |
| 19 | 46 | 32.6 | 457 | 1 PH4H_CAEEL | P09325 caenorhabdi |
| 20 | 45.5 | 32.3 | 432 | 1 SYWC_YEAST | Q12109 saccharomyc |
| 21 | 45 | 31.9 | 452 | 1 CN17_DICDI | P12019 dictyosteli |
| 22 | 44 | 31.2 | 172 | 1 Y427_UREUR | Q56564 ureaplasma |
| 23 | 44 | 31.2 | 371 | 1 CYB_ELANI | Q9mlk8 elapsosidea |
| 24 | 44 | 31.2 | 664 | 1 MS16_YEAST | P15424 saccharomyc |
| 25 | 44 | 31.2 | 772 | 1 C1PB_CLOTM | Q01866 clostridium |
| 26 | 44 | 31.2 | 793 | 1 MUTS_THEMA | P74926 thermotoga |
| 27 | 44 | 31.2 | 841 | 1 MYFC_YEREN | P33408 yersinia en |
| 28 | 44 | 31.2 | 1853 | 1 C1PA_CLOTM | Q06851 clostridium |
| 29 | 43.5 | 30.9 | 288 | 1 T2D2_STRPN | P09357 streptococc |
| 30 | 43 | 30.5 | 145 | 1 HA17_CLOBO | P46083 clostridium |
| 31 | 43 | 30.5 | 282 | 1 PANC_AQUAE | O67891 aquifex aeo |
| 32 | 43 | 30.5 | 370 | 1 CYB_MICKI | Q9mlk2 micropechis |
| 33 | 43 | 30.5 | 371 | 1 CYB_LOXBI | O48100 loxocemus b |

RESULT 1
CA34_HUMAN STANDARD; PRT; 1670 AA.
ID CA34_HUMAN
AC Q01955; Q9BQT2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Redders S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues";
RL J. Biol. Chem. 269:23013-23017(1994).
RN [2]
RP REVISIONS.
RA Leinonen A.;
RX Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
RP CVS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
RP PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064696; PubMed=1134255;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
RN [4]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:19780-19784(1992).
RN [5]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1882840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Redders S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RX TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;

ALIGNMENTS

| | | | | | | |
|----|----|------|------|---|------------|--------------------|
| 34 | 43 | 30.5 | 397 | 1 | YMP7_YEAST | Q04359 saccharomyc |
| 35 | 43 | 30.5 | 642 | 1 | YB9W_YEAST | P38352 saccharomyc |
| 36 | 43 | 30.5 | 680 | 1 | PBP2_STRPN | P10524 streptococc |
| 37 | 43 | 30.5 | 1462 | 1 | NCO2_MOUSE | O61026 mus musculu |
| 38 | 43 | 30.5 | 1464 | 1 | NCO2_HUMAN | Q15596 homo sapien |
| 39 | 43 | 30.5 | 1465 | 1 | NCO2_RAT | Q9wll9 rattus norv |
| 40 | 42 | 29.8 | 286 | 1 | VNS4_RSVN | P00847 rice stripe |
| 41 | 42 | 29.8 | 286 | 1 | VNS4_RSVT | P34961 rice stripe |
| 42 | 42 | 29.8 | 370 | 1 | CYB_BUNFA | Q9mlj8 bungarus fa |
| 43 | 42 | 29.8 | 385 | 1 | CHEE_BORBU | Q45047 borrelia bu |
| 44 | 42 | 29.8 | 385 | 1 | GUNF_FUSOX | P48239 fusarium ox |
| 45 | 42 | 29.8 | 417 | 1 | YGA4_YEAST | P53196 saccharomyc |

Mon Apr 5 07:53:14 2004

RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.; cloning of the human Goodpasture antigen demonstrates it
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601(1992).
RN [7].
RP SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=Kidney;
RA Ding J.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DDBJ databases.
RN [8].
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NCI domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [9].
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36";
RL FEBS Lett. 424:11-16(1998).
RN [10].
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [11].
RP VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7987301;
RA Lemink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monens H.A.H., van Oost B.A., Brunner H.G.,
RA Readers S.T., Smeets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT recessive Alport syndrome";
RL Hum. Mol. Genet. 3:1269-1273(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist. Isoforms differ in
CC the C-terminal part of the NCI domain;
CC Name=1;
CC IsoId=Q01955-1; Sequence=Displayed;
CC Name=2; Synonyms=V;
CC IsoId=Q01955-2; Sequence=VSP_001170;
CC Name=3; Synonyms=I5;
CC IsoId=Q01955-3; Sequence=VSP_001171;
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
CC -!- DISBASE: Antibodies against the NCI domain of alpha3(IV) mediate
CC the autoimmune disease Goodpasture syndrome [MIM:233450], which is
CC characterized by hematuria and pulmonary hemorrhage.
CC -!- DISBASE: Defects in COL4A3 are a cause of autosomal recessive
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder
CC characterized by progressive glomerulonephritis, renal failure,
CC hematuria, ocular abnormalities and deafness. The recessive form
CC occurs equally between males and females.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X80031; CAA56335.1; -!
CC EMBL; AJ288487; CAC36101.1; JOINED.
CC EMBL; AJ288488; CAC36101.1; JOINED.
CC EMBL; AJ288489; CAC36101.1; JOINED.
CC EMBL; AJ288490; CAC36101.1; JOINED.
CC EMBL; AJ288491; CAC36101.1; JOINED.
CC EMBL; AJ288492; CAC36101.1; JOINED.
CC EMBL; AJ288493; CAC36101.1; JOINED.
CC EMBL; AJ288494; CAC36101.1; JOINED.
CC EMBL; AJ288495; CAC36101.1; JOINED.
CC EMBL; AJ288496; CAC36101.1; JOINED.
CC EMBL; AJ288497; CAC36101.1; JOINED.
CC EMBL; AJ288498; CAC36101.1; JOINED.
CC EMBL; AJ288499; CAC36101.1; JOINED.
CC EMBL; AJ288500; CAC36101.1; JOINED.
CC EMBL; AJ288501; CAC36101.1; JOINED.
CC EMBL; AJ288502; CAC36101.1; JOINED.
CC EMBL; AJ288503; CAC36101.1; JOINED.
CC EMBL; AJ288504; CAC36101.1; JOINED.
CC EMBL; AJ288505; CAC36101.1; JOINED.
CC EMBL; AJ288506; CAC36101.1; JOINED.
CC EMBL; AJ288507; CAC36101.1; JOINED.
CC EMBL; AJ288508; CAC36101.1; JOINED.
CC EMBL; AJ288509; CAC36101.1; JOINED.
CC EMBL; AJ288510; CAC36101.1; JOINED.
CC EMBL; AJ288511; CAC36101.1; JOINED.
CC EMBL; AJ288512; CAC36101.1; JOINED.
CC EMBL; AJ288513; CAC36101.1; JOINED.
CC EMBL; AJ288514; CAC36101.1; JOINED.
CC EMBL; AJ288515; CAC36101.1; JOINED.
CC EMBL; AJ288516; CAC36101.1; JOINED.
CC EMBL; AJ288517; CAC36101.1; JOINED.
CC EMBL; AJ288518; CAC36101.1; JOINED.
CC EMBL; AJ288519; CAC36101.1; JOINED.
CC EMBL; AJ288520; CAC36101.1; JOINED.
CC EMBL; AJ288521; CAC36101.1; JOINED.
CC EMBL; AJ288522; CAC36101.1; JOINED.
CC EMBL; AJ288523; CAC36101.1; JOINED.
CC EMBL; AJ288524; CAC36101.1; JOINED.
CC EMBL; AJ288525; CAC36101.1; JOINED.
CC EMBL; AJ288526; CAC36101.1; JOINED.
CC EMBL; AJ288527; CAC36101.1; JOINED.
CC EMBL; AJ288528; CAC36101.1; JOINED.
CC EMBL; AJ288529; CAC36101.1; JOINED.
CC EMBL; AJ288530; CAC36101.1; JOINED.
CC EMBL; AJ288531; CAC36101.1; JOINED.
CC EMBL; AJ288532; CAC36101.1; JOINED.
CC EMBL; AJ288533; CAC36101.1; JOINED.
CC EMBL; AJ288534; CAC36101.1; JOINED.
CC EMBL; AJ288535; CAC36101.1; JOINED.
CC EMBL; AJ288536; CAC36101.1; JOINED.

```

Query Match      89.4%; Score 126; DB 1; Length 1670;
Best Local Similarity 92.3%; Pred. No. 1.3e-10;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
DB 1495 QRFTTTPFLFCNVNDVCFNFSRNDYS 1520

RESULT 2
CA34 BOVIN STANDARD; PRT; 471 AA.
ID CA34 BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]_TaxID=9913;
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Reiders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RT encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type

```

```

IV collagens.
-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M63139; AA62708.1; -.
PIR; A39024; A39024.
InterPro; IPR008160; Collagen.
InterPro; IPR001442; Procollagen_C.
PFam; PF01413; C4; 2.
PFam; PF01391; Collagen; 4.
ProDom; PD003923; ProcollagenC4; 1.
SMART; SM00111; C4; 2.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 1
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT DOMAIN 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match      88.7%; Score 125; DB 1; Length 471;
Best Local Similarity 88.5%; Pred. No. 4.4e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
DB 296 QRFTTTPFLFCNVNDVCFNFSRNDYS 321

RESULT 3
CAS3 CANFA STANDARD; PRT; 754 AA.
ID CAS3 CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; TISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thorner P.S., Marrano P., Baumal R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV.";
RN [2]
RP Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-

```

alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

-!- SUBCELLULAR LOCATION: Cell surface (potential).

-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

-!- PM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

-!- PM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

-!- DISEASE: A defect in COL4A5 has been found to be the cause of canine X-linked hereditary nephritis (HN), a disease similar to that in humans (also referred to as Alport syndrome) characterized by progressive renal failure and neurosensory deafness.

-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U07888; AAB60258.1; -.

PIR: A55267; A55267.

InterPro: IPR008161; C1g helix.

InterPro: IPR008160; Collagen.

InterPro: IPR001442; Procollagen4_C.

Pfam: PF01413; C4; 2.

Pfam: PF01391; Collagen; 8.

ProDom: PDOC00007; C1g helix; 1.

ProDom: PDOC03923; ProcollagenC4; 1.

SMART: SM00111; C4; 2.

Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Cell adhesion.

NON_TER 1

DOMAIN <1 530 TRIPLE-HELICAL REGION.

531 >754 NON-HELICAL REGION (NC1).

DISULFID 552 643 OR 640 (BY SIMILARITY).

DISULFID 595 640 OR 643 (BY SIMILARITY).

DISULFID 597 603 BY SIMILARITY.

DISULFID 662 ? OR 754 (BY SIMILARITY).

DISULFID 696 754 BY SIMILARITY.

DISULFID 708 714 BY SIMILARITY.

NON_TER 754 754

SEQUENCE 754 AA; 73537 MW; DSE321C287FA925B CRC64;

Query Match 78.0%; Score 110; DB 1; Length 754;

Best Local Similarity 73.1%; Pred. No. 1.3e-08;

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

2 QRFETMPLEFSNVNDVSNFASNDYS 27

587 RRFSTMPFMFCNNVNCVSNFASNDYS 612

RESULT 4

CAS4_HUMAN

AC P29400; Q16005; Q16126; PRT; 1685 AA.

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 5(IV) chain precursor.

GN COL4A5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94165049; PubMed=8120014;

RA Zhou J., Leinonen A., Tryggvason K.;

RT "Structure of the human type IV collagen COL4A5 gene.";

RN [2]

RJ J. Biol. Chem. 269:6608-6614(1994).

RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.

RC TISSUE=Kidney;

RX MEDLINE=923115923; PubMed=1352287;

RA Zhou J., Herrz J.M., Leinonen A., Tryggvason K.;

RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";

RN [3]

RJ J. Biol. Chem. 267:12475-12481(1992).

RP SEQUENCE OF 85-1685 FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=90337990; PubMed=2380186;

RA Pihlajaniemi T., Pihlajainen E.R., Myers J.C.;

RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";

RN [4]

RJ J. Biol. Chem. 265:13759-13766(1990).

RP SEQUENCE OF 924-1685 FROM N.A.

RX MEDLINE=91169491; PubMed=2004755;

RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;

RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";

RN [5]

RJ Genomics 9:1-9(1991).

RP SEQUENCE OF 914-1685 FROM N.A.

RX MEDLINE=90160375; PubMed=1689491;

RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;

RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the RT locus of X chromosome-linked Alport syndrome.";

RN [6]

RJ Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).

RP SEQUENCE OF 1442-1471 FROM N.A.

RX MEDLINE=90252791; PubMed=2339699;

RA Myers J.C., Jones T.A., Pihlajainen E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;

RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";

RN [7]

RJ Am. J. Hum. Genet. 46:1024-1033(1990).

RP SEQUENCE OF 1-20 FROM N.A.

RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;

RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.

RN [8]

RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).

RX MEDLINE=94133540; PubMed=8301933;

RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;

RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";

RN [9]

RJ Kidney Int. 44:1316-1321(1993).

RP REVIEW ON VARIANTS.

RX MEDLINE=97338662; PubMed=9195222;

RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;

RT "The clinical spectrum of type IV collagen mutations.";

RN [10]

RJ Hum. Mutat. 9:477-499(1997).

RP VARIANT AS SER-1564.

RX MEDLINE=91169492; PubMed=1672282;
 RA Zhou J., Barker D.F., Hosick S.L., Gregory M.C., Atkin C.L.,
 RA Tryggvason K.;
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
 RT conserved cysteine to serine in Alport syndrome.";
 RL Genomics 9:10-18(1991).
 RN [11]
 RP VARIANT AS ARG-325.
 RX MEDLINE=92303559; PubMed=1376965;
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
 RA Tryggvason K., Gubler M.-C., Antignac C.;
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5
 RT (IV) chain associated with X-linked Alport syndrome: characterization
 RT of the mutation by direct sequencing of PCR-amplified lymphoblast
 RT cDNA fragments.";
 RL Am. J. Hum. Genet. 51:135-142(1992).
 RN [12]
 RP VARIANT AS GLU-325.
 RX MEDLINE=93244772; PubMed=1363780;
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
 RA Rizzoni G.F., de Marchi M.;
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to
 RT glutamic acid in Alport syndrome.";
 RL Hum. Mol. Genet. 1:127-129(1992).
 RN [13]
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
 RX MEDLINE=94010948; PubMed=8406498;
 RA Lemnick H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
 RA Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W.,
 RA van Oost B.A., Smeets H.J.M.;
 RT "Identification of four novel mutations in the COL4A5 gene of
 RT patients with Alport syndrome.";
 RL Genomics 17:485-489(1993).
 RN [14]
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;
 RP ARG-869; ARG-872 AND CYS-1241.
 RX MEDLINE=96213750; PubMed=8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RA Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RT common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE=96213754; PubMed=8651296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RT exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
 RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RT syndrome.";
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RA Tverskaya S., Bobryna V., Tsalykova F., Ignatova M.,

RA Krasnopol'skaya X., Evgrafov O.;
 RT "Substitution of Ala98D in noncollagen domain of alpha 5(IV) collagen
 RT chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RN [19]
 RP VARIANT AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RT syndrome and COL4A5 R1677Q.";
 RL Hum. Genet. 99:681-684(1997).
 RN [20]
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
 RP AND ASP-1596.
 RX MEDLINE=98112435; PubMed=9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 RA Trivelli A., Pignatelli A.R., Ragaio M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 RT Alport syndrome.";
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).
 RN [21]
 RP VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;
 RP 802-GLY--PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
 RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
 RX MEDLINE=99063529; PubMed=9948783;
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,
 RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected
 RT Alport syndrome using PCR and direct DNA sequencing.";
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
 RN [22]
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
 RX MEDLINE=20030197; PubMed=10561141;
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
 RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
 RT "Detection of mutations in the COL4A5 gene in over 90% of male
 RT patients with X-linked Alport's syndrome by RT-PCR and direct
 RT sequencing.";
 RL Am. J. Kidney Dis. 34:854-862(1999).
 RN [23]
 RP VARIANT AS ARG-822.
 RX MEDLINE=96213750; PubMed=8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RA Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RT common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RP VARIANTS AS.
 RX MEDLINE=96213754; PubMed=8651296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RT exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
 RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RT syndrome.";
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RA Tverskaya S., Bobryna V., Tsalykova F., Ignatova M.,

Query Match 78.0%; Score 110; DB 1; Length 1685;
 Best Local Similarity 73.1%; Pred No. 3.2e-08;
 Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27

DB 1511 RRFSTWPFMFCNNVNCVNFASRNDYS 1536

RESULT 5

CA14 HUMAN

ID CA14 HUMAN STANDARD; PRT; 1669 AA.

AC P02462;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 1(IV) chain precursor.

GN COL4A1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89340433; PubMed=2701944;

RA Soiminen R., Huctari M., Ganguly A., Prockop D.J., Tryggvason K.;

RT "Structural organization of the gene for the alpha 1 chain of human

RT type IV collagen.";

RL J. Biol. Chem. 264:13565-13571(1989).
RN [2]
RP SEQUENCE OF 46-1257 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88083584; PubMed=3691802;
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
RT "Complete primary structure of the alpha 1-chain of human basement
RT membrane (type IV) collagen.";
RL FEBS Lett. 225:188-194(1987).
RN [3]
RP SEQUENCE OF 1-943 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89029471; PubMed=3311751;
RA Brazel D., Oberbaumer I., Dieringer H., Babel W., Glanville R.W.,
RA Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
RT basement membrane collagen (type IV) reveals 21 non-triplet
RT interruptions located within the collagenous domain.";
RL Eur. J. Biochem. 168:529-536(1987).
RN [4]
RP SEQUENCE OF 28-243.
RX MEDLINE=86004708; PubMed=4043082;
RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
RT "Amino acid sequence of the N-terminal aggregation and cross-linking
RT region (7S domain) of the alpha 1 (IV) chain of human basement
RT membrane collagen.";
RL Eur. J. Biochem. 152:213-219(1985).
RN [5]
RP SEQUENCE OF 534-1447.
RX MEDLINE=85003629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RT "Structure of human-basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1(IV) chain.";
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85207819; PubMed=2581969;
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA Chung M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Keralides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Lysines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26576; AAA53098.1; JOINED.
CC EMBL; J04217; AAA53098.1; JOINED.
CC EMBL; M26550; AAA53098.1; JOINED.
CC EMBL; M26540; AAA53098.1; JOINED.
CC EMBL; M26542; AAA53098.1; JOINED.
CC EMBL; M26543; AAA53098.1; JOINED.
CC EMBL; M26544; AAA53098.1; JOINED.
CC EMBL; M26545; AAA53098.1; JOINED.
CC EMBL; M26546; AAA53098.1; JOINED.
CC EMBL; M26547; AAA53098.1; JOINED.
CC EMBL; M26537; AAA53098.1; JOINED.
CC EMBL; M26538; AAA53098.1; JOINED.
CC EMBL; M26548; AAA53098.1; JOINED.
CC EMBL; M26549; AAA53098.1; JOINED.
CC EMBL; M26551; AAA53098.1; JOINED.
CC EMBL; M26552; AAA53098.1; JOINED.
CC EMBL; M26553; AAA53098.1; JOINED.
CC EMBL; M26554; AAA53098.1; JOINED.
CC EMBL; M26555; AAA53098.1; JOINED.
CC EMBL; M26556; AAA53098.1; JOINED.
CC EMBL; M26557; AAA53098.1; JOINED.
CC EMBL; M26558; AAA53098.1; JOINED.
CC EMBL; M26559; AAA53098.1; JOINED.
CC EMBL; M26560; AAA53098.1; JOINED.
CC EMBL; M26561; AAA53098.1; JOINED.
CC EMBL; M26562; AAA53098.1; JOINED.
CC EMBL; M26563; AAA53098.1; JOINED.
CC EMBL; M26564; AAA53098.1; JOINED.
CC EMBL; M26565; AAA53098.1; JOINED.
CC EMBL; M26566; AAA53098.1; JOINED.
CC EMBL; M26567; AAA53098.1; JOINED.
CC EMBL; M26568; AAA53098.1; JOINED.
CC EMBL; M26569; AAA53098.1; JOINED.
CC EMBL; M26570; AAA53098.1; JOINED.
CC EMBL; M26571; AAA53098.1; JOINED.
CC EMBL; M26572; AAA53098.1; JOINED.
CC EMBL; M26573; AAA53098.1; JOINED.
CC EMBL; M26574; AAA53098.1; JOINED.
CC EMBL; M26575; AAA53098.1; JOINED.
CC EMBL; Y00706; CAA68698.1; -
CC EMBL; X05561; CAA29075.1; -
CC EMBL; M10940; AAA52006.1; -
CC EMBL; M11315; AAA52042.1; -
CC PIR; S16876; CGHU4B.
CC Genew; HGNC:2202; COL4A1.

DR MIM; 120130; --
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; Clg_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT CARBOHYD 126 126
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1562.
FT DISULFID 1570 1665 OR 1665.
FT DISULFID 1604 1662
FT DISULFID 1616 1622
FT CONFLICT 237 238 SG -> XE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 E -> Q (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9BA84 CRC64;

Query Match 77.3%; Score 109; DB 1; Length 1669;
Best Local Similarity 73.1%; Pred. No. 4.4e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFVTMPFLPSNVNDVSNFASRNDYS 27
:::|||||:|:|||||
Db 1495 RKFTMPFLFCNNVNCNFCASRNDYS 1520

RESULT 6
CAL4_MOUSE
ID CAL4_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=89112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen

RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RT Vogeli G., Voss T., Siebold B., Gnanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes F., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J03758; AAA37439.1; -;
 CC EMBL; M23333; AAA51625.1; -;
 CC EMBL; J04694; AAA50292.1; -;
 CC EMBL; X06777; CAA29946.1; -;
 CC EMBL; X02201; CAA26132.1; -;
 CC EMBL; M15832; AAA37340.1; -;
 CC EMBL; M14042; AAA37342.1; -;
 CC EMBL; M12879; AAA37343.1; -;
 CC EMBL; M13024; -; NOT_ANNOTATED_CDS.
 CC EMBL; M13025; -; NOT_ANNOTATED_CDS.
 CC EMBL; M13026; AAA37344.1; -;
 CC EMBL; M13027; AAA37345.1; -;
 CC EMBL; M13043; AAA37346.1; -;
 CC EMBL; J04448; AAA37437.1; -;
 CC EMBL; J03525; CGMS4B.
 CC MGD; MGI:88454; Col4a1.
 CC GO; GO:0005604; C:Basement membrane; IDA.
 CC InterPro; IPR008161; Clg_helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR001442; Procollagn4_C.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 23.
 CC ProDom; PD000007; Clg_helix; 6.
 CC ProDom; PD003923; Procollagn4; 1.
 CC SMART; SM00111; C4; 2.
 CC Repeat; Hydroxylation; Connective tissue; Basement membrane;
 CC Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 CC SIGNAL 1 27
 CC PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 CC CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 CC DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 CC DOMAIN 1441 1669 NONHELICAL REGION (NC1).
 CC DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
 CC DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 CC DISULFID 1505 1511 BY SIMILARITY.
 CC DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
 CC DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
 CC DISULFID 1616 1622 BY SIMILARITY.
 CC CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CONFLICT 26 26 A -> P (IN REF. 2).
 CC CONFLICT 186 186 S -> L (IN REF. 2).
 CC CONFLICT 319 319 Q -> S (IN REF. 2).
 CC CONFLICT 369 369 L -> L (IN REF. 2).
 CC CONFLICT 403 403 P -> F (IN REF. 2).
 CC CONFLICT 481 481 P -> L (IN REF. 2).
 CC CONFLICT 493 493 S -> I (IN REF. 2).
 CC CONFLICT 712 712 E -> Q (IN REF. 2).
 CC CONFLICT 813 813 E -> Q (IN REF. 2).
 CC CONFLICT 982 982 Q -> H (IN REF. 2).
 CC CONFLICT 1397 1397 V -> S (IN REF. 3).
 CC SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;
 CC
 CC Query Match 77.3%; Score 109; DB 1; Length 1669;
 CC Best Local Similarity 73.1%; Pred. No. 4.4e-08;
 CC Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 2 QRRTMPFLFSNVNVDNFSASNDYS 27
 CC Db 1495 RKFTMPFLFCNNVNCNFSASNDYS 1520
 CC
 CC RESULT 7
 CC CA24_ASCSU STANDARD; PRT; 1763 AA.
 CC ID CA24_ASCSU
 CC AC P27393;
 CC DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.
 OC NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS I AND II).
 RX MEDLINE=91340768; PubMed=1714907;
 RA Pettitt J., Kingston I.B.;
 RT "The complete primary structure of a nematode alpha 2(IV) collagen
 and the partial structural organization of its gene.";
 RL J. Biol. Chem. 266:16149-16156(1991).
 CC !- FUNCTION: Collagens type IV is specific for basement membranes.
 CC !- SURUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
 CC Type IV collagen forms a mesh-like network linked through
 CC intermolecular interactions between 7S domains and between NC1
 CC domains.
 CC !- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=I;
 CC IsoId=P27393-1; Sequence=Displayed;
 CC Name=II;
 CC IsoId=P27393-2; Sequence=VSP_001159;
 CC !- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the
 CC G-X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC !- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC !- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67507; AAAL8014.1; -;
 CC PIR; S16366; S16366.
 CC InterPro; IPR008161; Clg_helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR001442; Procollagn4_C.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 25.
 CC ProDom; PD000007; Clg_helix; 6.
 CC ProDom; PD003923; Procollagn4; 1.
 CC SMART; SM00111; C4; 2.
 CC Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 CC Alternative splicing; Glycoprotein; Signal.
 CC SIGNAL 1 26 POTENTIAL.
 CC CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.
 CC DOMAIN 27 42 7S DOMAIN.
 CC DOMAIN 43 1529 TRIPLE-HELICAL REGION.
 CC DOMAIN 1530 1763 NONHELICAL REGION (NC1).
 CC DISULFID 1548 1637 OR 1634 (BY SIMILARITY).
 CC DISULFID 1581 1634 OR 1637 (BY SIMILARITY).
 CC DISULFID 1593 1599 BY SIMILARITY.
 CC DISULFID 1656 1752 OR 1749 (BY SIMILARITY).
 CC DISULFID 1690 1749 OR 1752 (BY SIMILARITY).
 CC DISULFID 1702 1709 BY SIMILARITY.
 CC CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 249 249 O-LINKED (XYL. . .) (GLYCOSAMINOGLYCAN
 CC (IN ISOFORM II) (POTENTIAL).
 CC VARSPLIC 230 266 GEQGPGRGPGPPGPGTGTGAKTGTGEGAPGMKGEK ->
 CC GLIGPAGPGPGPGPREFTGSGIVGRHSGDKGVK (in

Matches 17; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFTHMPFLPSNVNDVSNFASRNDYS 27
DB 1581 QRFTHMPFLPSNVNDVSNFASRNDYS 1606

RESULT 9
CAL14_CAEEL STANDARD; PRT; 1758 AA.
ID CAL14_CAEEL
AC P17139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV), chain precursor.
GN EMB-9 OR CIB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peliceridae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirksen J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimer of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
intermolecular interactions between 7S domains and between NC1
domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type

CC IV collagens.
CC -!- DISEASE: Mutations in this gene cause temperature-sensitive
lethality during late embryogenesis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; X56979; CAA40299.1; -;
DR EMBL; Z27078; CAA81584.3; -;
DR EMBL; J05067; AAB59179.1; -;
DR PIR; S40991; S40991.
DR WormPep; K04H4.1; C532462.
DR InterPro; IPR008161; Cig_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 27.
DR ProDom; PD000007; Cig_helix; 11.
DR ProDom; PD003923; Procollagen4; 1.
DR SMART; SM00111; C4; 2.
DR Repeat; Hydroxylation; Connective tissue; Basement membrane;
KW Extracellular matrix; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 20
FT PROPEP 21 2194 FT AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN ?195 1758 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 195 1529 TRIPLE-HELICAL REGION.
FT DISULFID 1530 1758 NONHELICAL REGION (NC1).
FT DISULFID 1549 1640 OR 1637 (BY SIMILARITY).
FT DISULFID 1552 1637 OR 1640 (BY SIMILARITY).
FT DISULFID 1594 1600 BY SIMILARITY.
FT DISULFID 1659 1754 OR 1751 (BY SIMILARITY).
FT DISULFID 1693 1751 OR 1754 (BY SIMILARITY).
FT DISULFID 1705 1711 BY SIMILARITY.
FT VARIANT 402 402 G -> E (IN MUTANT G34).
FT VARIANT 408 408 G -> E (IN MUTANT G23/HCT0).
FT CONFLICT 302 391 LDNGKRGDGVPGNGKSGQSGQGLGTPGYPTKGAGE
LPYGRPGEGDCGPGEGTGEAGPHGAGFGVGGGK
GLPGHDLG -> AGORVSCPNKKLFLFCRVNTEFOE
IMKEDPKENKDLLELODTQLREGLNQDTQEDQVSKETV
DKQHLKELVRLDMELKDSFEKAKCQDMVVS (IN
REF. 2).
FT CONFLICT 581 581 G -> R (IN REF. 2).
FT CONFLICT 768 768 P -> R (IN REF. 2).
FT CONFLICT 830 830 D -> V (IN REF. 2).
FT CONFLICT 1514 1514 P -> Q (IN REF. 4).
FT CONFLICT 1722 1722 P -> L (IN REF. 2).
SQ SEQUENCE 1758 AA; 170857 MW; 7083D9AF63E05D45 CRC64;

Query Match 61.7%; Score 87; DB 1; Length 1758;
Best Local Similarity 60.0%; Pred. NO. 8.8e-05;
Matches 15; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 RTTHMPFLPSNVNDVSNFASRNDYS 27
DB 1585 KFTMTPEFPCNNVNSVCHVSSRNDYS 1609

RESULT 10
CA64_HUMAN STANDARD; PRT; 1691 AA.
ID CA64_HUMAN
AC Q14031; Q12823; Q14033; Q9NQMS; QNTX3; Q9UJ76; Q9UMG6; Q9Y4L4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eye, and Kidney;
RX MEDLINE=94171779; PubMed=8125972;
RA Ohashi T., Sugimoto M., Mattai M.-G., Ninomiya Y.;
RT identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RL J. Biol. Chem. 269:7520-7526(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Reeder S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Reeder S.T., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grafham D., Lawlor S., Wilson S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=8356449;
RA Zhou J., Nishizaki T., Smeets H., Antignac C., Laurila P.,
de Paeppe A., Tryggvason K., Reeder S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169(1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM) forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCL) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCL domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

CC EMBL; D21337; BAA04809.1; -;
DR EMBL; U04845; AAA19569.2; -;
DR EMBL; U47004; AAB19038.1; JOINED.
DR EMBL; U46959; AAB19038.1; JOINED.
DR EMBL; U46961; AAB19038.1; JOINED.
DR EMBL; U46962; AAB19038.1; JOINED.
DR EMBL; U46963; AAB19038.1; JOINED.
DR EMBL; U46964; AAB19038.1; JOINED.
DR EMBL; U46965; AAB19038.1; JOINED.
DR EMBL; U46966; AAB19038.1; JOINED.
DR EMBL; U46967; AAB19038.1; JOINED.
DR EMBL; U46968; AAB19038.1; JOINED.
DR EMBL; U46969; AAB19038.1; JOINED.
DR EMBL; U46970; AAB19038.1; JOINED.
DR EMBL; U46971; AAB19038.1; JOINED.
DR EMBL; U46972; AAB19038.1; JOINED.
DR EMBL; U46973; AAB19038.1; JOINED.
DR EMBL; U46974; AAB19038.1; JOINED.
DR EMBL; U46975; AAB19038.1; JOINED.
DR EMBL; U46976; AAB19038.1; JOINED.
DR EMBL; U46977; AAB19038.1; JOINED.
DR EMBL; U46978; AAB19038.1; JOINED.
DR EMBL; U46979; AAB19038.1; JOINED.
DR EMBL; U46980; AAB19038.1; JOINED.
DR EMBL; U46981; AAB19038.1; JOINED.
DR EMBL; U46982; AAB19038.1; JOINED.
DR EMBL; U46983; AAB19038.1; JOINED.
DR EMBL; U46984; AAB19038.1; JOINED.
DR EMBL; U46985; AAB19038.1; JOINED.
DR EMBL; U46986; AAB19038.1; JOINED.
DR EMBL; U46987; AAB19038.1; JOINED.
DR EMBL; U46988; AAB19038.1; JOINED.
DR EMBL; U46989; AAB19038.1; JOINED.
DR EMBL; U46990; AAB19038.1; JOINED.
DR EMBL; U46991; AAB19038.1; JOINED.
DR EMBL; U46992; AAB19038.1; JOINED.
DR EMBL; U46993; AAB19038.1; JOINED.
DR EMBL; U46994; AAB19038.1; JOINED.
DR EMBL; U46995; AAB19038.1; JOINED.
DR EMBL; U46996; AAB19038.1; JOINED.
DR EMBL; U46997; AAB19038.1; JOINED.
DR EMBL; U46998; AAB19038.1; JOINED.
DR EMBL; U46999; AAB19038.1; JOINED.
DR EMBL; U47000; AAB19038.1; JOINED.
DR EMBL; U47001; AAB19038.1; JOINED.
DR EMBL; U47002; AAB19038.1; JOINED.
DR EMBL; U47003; AAB19038.1; JOINED.
DR EMBL; U47004; AAB19039.1; -;
DR EMBL; U46960; AAB19039.1; JOINED.
DR EMBL; U46961; AAB19039.1; JOINED.
DR EMBL; U46962; AAB19039.1; JOINED.
DR EMBL; U46963; AAB19039.1; JOINED.
DR EMBL; U46964; AAB19039.1; JOINED.
DR EMBL; U46965; AAB19039.1; JOINED.
DR EMBL; U46966; AAB19039.1; JOINED.
DR EMBL; U46967; AAB19039.1; JOINED.
DR EMBL; U46968; AAB19039.1; JOINED.
DR EMBL; U46969; AAB19039.1; JOINED.
DR EMBL; U46970; AAB19039.1; JOINED.
DR EMBL; U46971; AAB19039.1; JOINED.
DR EMBL; U46972; AAB19039.1; JOINED.
DR EMBL; U46973; AAB19039.1; JOINED.
DR EMBL; U46974; AAB19039.1; JOINED.
DR EMBL; U46975; AAB19039.1; JOINED.
DR EMBL; U46976; AAB19039.1; JOINED.
DR EMBL; U46977; AAB19039.1; JOINED.
DR EMBL; U46978; AAB19039.1; JOINED.
DR EMBL; U46979; AAB19039.1; JOINED.
DR EMBL; U46980; AAB19039.1; JOINED.
DR EMBL; U46981; AAB19039.1; JOINED.
DR EMBL; U46982; AAB19039.1; JOINED.
DR EMBL; U46983; AAB19039.1; JOINED.

DR EMBL; U45984; AAB19039.1; JOINED.
 DR EMBL; U45985; AAB19039.1; JOINED.
 DR EMBL; U45986; AAB19039.1; JOINED.
 DR EMBL; U45987; AAB19039.1; JOINED.
 DR EMBL; U45988; AAB19039.1; JOINED.
 DR EMBL; U45989; AAB19039.1; JOINED.
 DR EMBL; U45990; AAB19039.1; JOINED.
 DR EMBL; U45991; AAB19039.1; JOINED.
 DR EMBL; U45992; AAB19039.1; JOINED.
 DR EMBL; U45993; AAB19039.1; JOINED.
 DR EMBL; U45994; AAB19039.1; JOINED.
 DR EMBL; U45995; AAB19039.1; JOINED.
 DR EMBL; U45996; AAB19039.1; JOINED.
 DR EMBL; U45997; AAB19039.1; JOINED.
 DR EMBL; U45998; AAB19039.1; JOINED.
 DR EMBL; U45999; AAB19039.1; JOINED.
 DR EMBL; U47000; AAB19039.1; JOINED.
 DR EMBL; U47001; AAB19039.1; JOINED.
 DR EMBL; U47002; AAB19039.1; JOINED.
 DR EMBL; U47003; AAB19039.1; JOINED.
 DR EMBL; AL034369; CAA22265.1; -.
 DR EMBL; AL109943; CAB99263.1; -.
 DR EMBL; AL360880; CAB96748.1; -.
 DR EMBL; AL031177; CAA20120.1; -.
 DR EMBL; L22763; AAA16338.1; -.
 DR PIR; A54122; CGHU6B.
 DR Genew; HGNC:2208; COL4A6.
 DR MIM; 303631; -.
 DR GO; GO:0005587; C:collagen type IV; NAS.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
 DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
 DR InterPro; IPR008161; C1g helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 23.
 DR ProDom; PDOC00007; C1g_helix; 4.
 DR ProDom; PDOC3923; ProcollagenC4; 1.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
 KW Alternative splicing; Polymorphism.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
 FT DOMAIN 23 46 7S DOMAIN.
 Query Match 60.3%; Score 85; DB 1; Length 1691;
 Best Local Similarity 56.0%; Pred. NO. 0.00017;
 Matches 14; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
 OY 3 RPTMPLFSLNVDVNSFASNDYS 27
 DB 1518 RFTMPTFYCINVECHVARENDXS 1542
 RESULT 11
 C224 MOUSE
 ID CA24_MOUSE STANDARD; PRT; 1707 AA.
 AC P08122; Q61375;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 GN COL4A2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89197933; PubMed=2703491;
 RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
 RA Pihlajaniemi T., Kurkinen M.;

ET "The complete primary structure of mouse alpha 2(IV) collagen.
 RT Alignment with mouse alpha 1(IV) collagen.";
 RL J. Biol. Chem. 264:6318-6324 (1989).
 RN [2]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89066738; PubMed=3198626;
 RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
 RA "Head-to-head arrangement of murine type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277 (1988).
 RN [3]
 RP SEQUENCE OF 970-1480 FROM N.A.
 RX MEDLINE=86220192; PubMed=3011432;
 RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
 RA Deutzmann R., Timpl R., Kuehn K.;
 RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain.";
 RL Eur. J. Biochem. 157:49-56 (1986).
 RN [4]
 RP SEQUENCE OF 1480-1707 FROM N.A.
 RX MEDLINE=87054581; PubMed=3780963;
 RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
 RT "cDNA and protein sequence of the NCI domain of the alpha 2-chain of collagen IV and its comparison with alpha 1(IV).";
 RL FEBS Lett. 208:203-207 (1986).
 RN [5]
 RP SEQUENCE OF 1481-1707 FROM N.A.
 RX MEDLINE=87250460; PubMed=3597383;
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
 RA Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8496-8499 (1987).
 RN [6]
 RP SEQUENCE OF 1041-1489 FROM N.A.
 RX MEDLINE=87005245; PubMed=3758345;
 RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
 RT "Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";
 RL FEBS Lett. 206:29-32 (1986).
 RN [7]
 RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
 RX MEDLINE=85296379; PubMed=3839908;
 RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
 RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";
 RL Nature 317:177-179 (1985).
 RN [8]
 RP SEQUENCE OF 1-60 FROM N.A.
 RX MEDLINE=89071759; PubMed=3200851;
 RA Burdello P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).
 CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) - alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.
 CC -----

RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RL of basement-membrane type IV collagen.
 CC Eur. J. Biochem. 176:617-624 (1988).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 CC alpha 6(IV), each of which can form a triple helix structure
 CC with 2 other chains to generate type IV collagen network.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the
 CC G-X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X05562; CAA29076.1; -
 CC EMBL; X05610; CAA29098.1; -
 CC EMBL; J02760; AAA58422.1; -
 CC EMBL; M36963; AAA53099.1; -
 CC EMBL; X12784; CAA31275.1; -
 CC EMBL; J04217; AAA53097.1; -
 CC PIR; A32024; CGHU2B
 CC Genew; HGNC:2203; COL4A2.
 CC MIW; 120090; -
 CC GO; GO:0005587; C:collagen type IV; TAS.
 CC GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
 CC GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
 CC InterPro; IPR008161; C1q helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR001442; Procollag4_C.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 7.
 CC ProDom; PD000007; C1q helix; 7.
 CC ProDom; PD003923; ProcollagN4; 1.
 CC SMART; SM00111; C4; 2.
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Glycoprotein; Basement membrane; Collagen; Signal.
 CC SIGNAL 1 25
 CC PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 CC CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
 CC DOMAIN 184 1484 TRIPLE-HELICAL REGION.
 CC DOMAIN 1485 1712 NONHELICAL REGION (NC1).
 CC DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
 CC DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
 CC DISULFID 1549 1555 BY SIMILARITY.
 CC DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
 CC DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
 CC DISULFID 1658 1665 BY SIMILARITY.
 CC CARBOHYD 138 138 N-LINKED (GLCNAC. .).
 CC CONFLICT 471 471 R -> P (IN REF. 2).
 CC CONFLICT 683 683 A -> G (IN REF. 2).
 CC CONFLICT 1575 1575 M -> I (IN REF. 5).
 CC CONFLICT 1663 1663 G -> H (IN REF. 9).
 CC CONFLICT 1701 1701 H -> G (IN REF. 9).
 CC SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
 Query Match 58.2%; Score 82; DB 1; Length 1712;
 Best Local Similarity 64.0%; Pred. No. 0.00047;

Matches 16; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 3 RFTMPPLFSNVNDVSFASRNDYS 27
 DB 1540 RFTMPPLFCNGVGVYASRNDKS 1564
 RESULT 13
 CAA4. RABIT
 ID CAA4 RABIT STANDARD; PRT; 623 AA.
 AC F55787;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Collagen alpha 4(IV) chain (Fragment).
 GN COL4A4.
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CC NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Corneal endothelium;
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 RT alpha 4 chain of basement membrane collagen type IV and assignment of
 RT the gene to the distal long arm of human chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758(1992).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 CC alpha 6(IV), each of which can form a triple helix structure with
 CC 2 other chains to generate type IV collagen network.
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-
 CC X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J01477; -; NOT_ANNOTATED_CDS.
 CC FIR; A45137; A45137.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR001442; ProcollagN4_C.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 5.
 CC ProDom; PD003923; ProcollagN4; 1.
 CC SMART; SM00111; C4; 2.
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 CC NON_TER 1 1
 CC DOMAIN <1 392 TRIPLE-HELICAL REGION.
 CC DOMAIN 393 623 NONHELICAL REGION (NC1).
 CC DISULFID 413 502 OR 499 (BY SIMILARITY).
 CC DISULFID 446 499 OR 502 (BY SIMILARITY).

Smeets H.J.M., Readers S.T.:
- Identification of mutations in the alpha 3(IV) and alpha 4(IV)
collagen genes in autosomal recessive Alport syndrome.";
[7]
Nat. Genet. 8:77-82(1994).
RN
VARIANT FBH GLU-897.
RP
MEDLINE=96379660; PubMed=8787673;
RX
Lemlink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
RA
Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
RA "Benign familial hematuria due to mutation of the type IV collagen
RT alpha4 gene."
RL J. Clin. Invest. 98:1114-1118(1996).
RN [8]

VARIANTS AS AND VARIANTS
RX MEDLINE=99011253; PubMed=9792860;
RA
Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
RA Cochat P., Gruenfeld J.-P., Falcoix J.-B., Guibet M.C., Antignac C.;
RA "Determination of the genomic structure of the COL4A4 gene and of
RT novel mutations causing autosomal recessive Alport syndrome.";
RL Am. J. Hum. Genet. 63:1329-1340(1998).
CC -I- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -I- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -I- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -I- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -I- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -I- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -I- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -I- DISEASE: Defects in COL4A4 are a cause of autosomal recessive
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder
CC characterized by progressive glomerulonephritis, renal failure,
CC hematuria, ocular abnormalities and deafness. The recessive form
CC occurs equally between males and females.
CC -I- DISEASE: Defects in COL4A4 are a cause of familial benign
CC hematuria (FBH) [MIM:141200] or thin basement membrane disease.
CC FBH is characterized by persistent hematuria, an electron
CC microscopically detectable thin glomerular basement membrane (GBM)
CC and an autosomal dominant mode of inheritance. Renal function
CC remains normal. In children, differentiation between FBH and AS
CC can be difficult, because both disorders are manifested by
CC persistent hematuria and thin GBM at that age.
CC -I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
The European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X81053; CAAS6943.1;
DR ENBL; AB008496; BAA2065.1; --
DR ENBL; D17391; BAA04214.1; --
DR PIR; A55360; CGHU1B
DR Genew; HGNC:2206; COL4A4.
DR MIM; 120131; --
DR MIM; 141200; --
DR MIM; 203780; --

FT DISULFID 458 464 BY SIMILARITY.
FT OR 616 (BY SIMILARITY).
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SQ SEQUENCE 623 AA; 62393 MW; CCBC9BB31243FE82 CRC64;

Query Match 48.2%; Score 68; DB 1; Length 623;
Best Local Similarity 45.8%; Pred.No. 0.018;
Matches 11; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 4 FTTFPLFSNVNDVNDFNFRNDYS 27
| : | | | : | : | : | : | | | |
DB 450 FSTLPFCYNHQVCHYAQRNDKS 473

RESULT 14
CA44_HUMAN
ID CA44_HUMAN STANDARD; PRT; 1690 AA.
IC P53420;
DT 01-OCT-1996 (Rel. 34, Created)
DT DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Collagen alpha 4(IV) chain precursor.
GN COL4A4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Kidney;
RX MEDLINE=95014445; PubMed=7523402;
RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Readers S.T.;
RA "Complete primary structure of the human type IV collagen alpha 4(IV)
RT chain. Comparison with structure and expression of the other alpha
RT (IV) chains.";
RT J. Biol. Chem. 269:26172-26177(1994).
RN [2]
RN SEQUENCE OF 1-23 FROM N.A.
RP MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RA "Two genes, COL4A3 and COL4A4 coding for the human alphas(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RT FEBS Lett. 424:11-16(1998).
RN [3]
RN SEQUENCE OF 1219-1690 FROM N.A.
RP TISSUE=EYE;
RX MEDLINE=93374047; PubMed=8365481;
RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
RA "cDNA isolation and partial gene structure of the human alpha 4(IV)
RT collagen chain.";
RT FEBS Lett. 330:122-128(1993).
RN [4]
RN SEQUENCE OF 1407-1507 FROM N.A.
RP MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RA "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
RN [5]
RN REVIEW ON VARIANTS.
RP MEDLINE=97338662; PubMed=9195222;
RA Lemink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RA "The clinical spectrum of type IV collagen mutations.";
Hum. Mutat. 9:477-499(1997).
RN [6]
RN VARIANT AS SER-1201.
RX MEDLINE=95078927; PubMed=7987396;
RA Mochizuki T., Lemink H.H., Mariyama M., Antignac C., Gubler M.-C.,
RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,

[illegible]

EMBL; M23704; AAA28404.1; -.
EMBL; M96575; AAB59184.1; -.
EMBL; J02727; AAA28423.1; -.

DR EMBL; M28334; AAA28422.1; -;
DR EMBL; V00200; CAA23486.2; -;
DR PIR; A31893; A31893.
DR FlyBase; FBgn0000299; Cg25C.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD000007; Clg_helix; 9.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 ?
FT CHAIN ? 1775
FT DOMAIN ? 1544
FT DOMAIN 1545 1775
FT DISULFID 1569 1655
FT DISULFID 1599 1652
FT DISULFID 1611 1617
FT DISULFID 1674 1770
FT DISULFID 1708 1767
FT DISULFID 1720 1727
FT CARBOHYD 72 72
FT CONFLICT 948 948
FT CONFLICT 997 997
FT CONFLICT 1357 1357
FT CONFLICT 1360 1360
FT CONFLICT 1373 1373
FT CONFLICT 1496 1496
FT CONFLICT 1507 1511
FT CONFLICT 1529 1529
FT CONFLICT 1733 1733
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AE23149525CD CRC64;
AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
COLLAGEN ALPHA 1(IV) CHAIN.
TRIPLE-HELICAL REGION.
NONHELICAL REGION (NC1).
OR 1652 (BY SIMILARITY).
OR 1655 (BY SIMILARITY).
BY SIMILARITY.
OR 1767 (BY SIMILARITY).
OR 1770 (BY SIMILARITY).
BY SIMILARITY.
N-LINKED (GLCNAC. . .) (PROBABLE).
L -> S (IN REF. 6).
S -> T (IN REF. 6).
Q -> K (IN REF. 5).
T -> I (IN REF. 5).
L -> R (IN REF. 5).
ETGNV -> RAGOR (IN REF. 5).
E -> K (IN REF. 5).
M -> I (IN REF. 5).

Query Match 48.2%; Score 68; DB 1; Length 1775;
Best Local Similarity 56.5%; Pred.No. 0.06;
Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 RFTTPEPLFSNVNDVNSFRND 25
||:|:|
Db 1602 RFSTLPVLSGQNNVCNYSRND 1624

Search completed: April 5, 2004, 06:59:40
Job time : 4.39952 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 16.2131 Seconds
(without alignments)
525.440 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 KQFTTTPFLFSNVNDVSNFASNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mhc:
- 7: sp_mhc:
- 8: sp_organella:
- 9: sp_phage:
- 10: sp_plant:
- 11: sp_prodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_virus:
- 16: sp_bacteriap:
- 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 126 | 89.4 | 212 | 6 Q28512 | Q28512 macaca mula |
| 2 | 126 | 89.4 | 245 | 4 Q9NYC4 | Q9NYC4 homo sapien |
| 3 | 125 | 88.7 | 203 | 6 Q29032 | Q29032 sus scrofa |
| 4 | 125 | 88.7 | 203 | 6 Q28682 | Q28682 oryctolagus |
| 5 | 125 | 88.7 | 212 | 6 Q28587 | Q28587 ovis aries |
| 6 | 120 | 85.1 | 161 | 11 Q61430 | Q61430 mus musculus |
| 7 | 120 | 85.1 | 210 | 6 Q28273 | Q28273 canis famill |
| 8 | 120 | 85.1 | 246 | 11 Q61435 | Q61435 mus musculus |
| 9 | 120 | 85.1 | 1669 | 11 Q9QZS0 | Q9QZS0 mus musculus |
| 10 | 116 | 82.3 | 230 | 11 Q63122 | Q63122 rattus norv |
| 11 | 110 | 78.0 | 179 | 11 P70165 | P70165 mus musculus |
| 12 | 110 | 78.0 | 253 | 11 Q61436 | Q61436 mus musculus |
| 13 | 110 | 78.0 | 585 | 11 Q80V57 | Q80V57 mus musculus |
| 14 | 110 | 78.0 | 799 | 11 Q8BNS7 | Q8BNS7 mus musculus |
| 15 | 110 | 78.0 | 886 | 4 Q9NUE7 | Q9NUE7 homo sapien |
| 16 | 110 | 78.0 | 1684 | 6 Q8HYC1 | Q8HYC1 canis famill |

| | | | | | |
|----|-----|------|------|-----------|---------------------|
| 17 | 110 | 78.0 | 1688 | 6 Q86622 | Q86622 canis famill |
| 18 | 110 | 78.0 | 1691 | 11 Q9ESQ2 | Q9ESQ2 mus musculus |
| 19 | 109 | 77.3 | 225 | 6 Q28271 | Q28271 canis famill |
| 20 | 109 | 77.3 | 226 | 11 Q91Q8 | Q91Q8 mus musculus |
| 21 | 109 | 77.3 | 229 | 4 Q8NF88 | Q8NF88 homo sapien |
| 22 | 109 | 77.3 | 229 | 4 Q9NYC5 | Q9NYC5 homo sapien |
| 23 | 109 | 77.3 | 229 | 13 Q919K3 | Q919K3 gallus gall |
| 24 | 109 | 77.3 | 1075 | 4 Q86X41 | Q86X41 homo sapien |
| 25 | 109 | 77.3 | 1621 | 4 Q9H4R9 | Q9H4R9 homo sapien |
| 26 | 100 | 70.9 | 1747 | 5 Q26640 | Q26640 strongyloce |
| 27 | 100 | 70.9 | 1752 | 5 Q07265 | Q07265 strongyloce |
| 28 | 88 | 62.4 | 1802 | 5 Q17163 | Q17163 brugia mala |
| 29 | 85 | 60.3 | 205 | 6 Q28274 | Q28274 canis famill |
| 30 | 85 | 60.3 | 546 | 11 Q9KX97 | Q9KX97 mus musculus |
| 31 | 85 | 60.3 | 1600 | 4 Q9UEH6 | Q9UEH6 homo sapien |
| 32 | 85 | 60.3 | 1691 | 11 Q9ESQ1 | Q9ESQ1 mus musculus |
| 33 | 82 | 58.2 | 202 | 6 Q28272 | Q28272 canis famill |
| 34 | 82 | 58.2 | 358 | 11 Q91VI3 | Q91VI3 mus musculus |
| 35 | 82 | 58.2 | 673 | 4 Q14052 | Q14052 homo sapien |
| 36 | 79 | 56.0 | 1723 | 5 Q9GQB1 | Q9GQB1 hydra atten |
| 37 | 69 | 48.9 | 1761 | 5 Q18407 | Q18407 drosophila |
| 38 | 69 | 48.9 | 1940 | 5 Q3VMV5 | Q3VMV5 drosophila |
| 39 | 68 | 48.2 | 312 | 11 Q64457 | Q64457 mus musculus |
| 40 | 68 | 48.2 | 1682 | 11 Q9QZR9 | Q9QZR9 mus musculus |
| 41 | 68 | 48.2 | 1779 | 5 Q9VMV4 | Q9VMV4 drosophila |
| 42 | 67 | 47.5 | 208 | 6 Q29468 | Q29468 canis famill |
| 43 | 67 | 47.5 | 1024 | 5 Q8T794 | Q8T794 anopheles g |
| 44 | 64 | 45.4 | 713 | 5 Q5GV24 | Q5GV24 sarcophaga |
| 45 | 59 | 41.8 | 1004 | 10 Q9LGN1 | Q9LGN1 cryza eativ |

ALIGNMENTS

RESULT 1

| | | | |
|--------|---|---------|-----------------------------------|
| Q28512 | PRELIMINARY; | PRT; | 212 AA. |
| ID | Q28512 | | |
| AC | Q28512; | | |
| DT | 01-NOV-1996 (TRENBLrel. 01, Created) | | |
| DT | 01-NOV-1996 (TRENBLrel. 01, Last sequence update) | | |
| DT | 01-OCT-2003 (TRENBLrel. 25, Last annotation update) | | |
| DE | Alpha-3 type IV collagen (Fragment). | | |
| GN | COL4A3. | | |
| OS | Macaca mulatta (Rhesus macaque). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae; | | |
| OC | Cercopitheidae; Macaca. | | |
| OX | NCBI_TaxID=9544; | | |
| RN | [1]_TaxID=9544; | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=Kidney cortex; | | |
| RA | Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I., | | |
| RA | Mason P.J., Pusey C.D.; | | |
| RT | "Properties and sequences of the Goodpasture antigen of different | | |
| RT | mammals." | | |
| RL | Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; I47280; AA91861.1; - | | |
| DR | GO; GO:000581; C:collagen; IEA. | | |
| DR | GO; GO:0005201; F:extracellular matrix structural constituent; IEA. | | |
| DR | GO; GO:0003676; F:nucleic acid binding; IEA. | | |
| DR | InterPro; IPR001442; Procollagn4_C. | | |
| DR | InterPro; IPR000504; RNA_rec_mot. | | |
| DR | Pfam; PF01413; C4; 2. | | |
| DR | ProDom; PD003923; ProcollagnC4; 1. | | |
| DR | SMART; SM00111; C4; 2. | | |
| DR | PROSITE; PS00030; RRM_RNP_1; 1. | | |
| KW | Collagen. | | |
| FT | NON_TER | 1 | 212 |
| FT | NON_TER | 212 | 212 |
| SQ | SEQUENCE | 212 AA; | 23469 MW; 4BC574A64E357E64 CRC64; |

Query Match

Best Local Similarity

89.4%; Score 126; DB 6; Length 212;
92.3%; Pred. No. 7.4e-11;

```
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27
    |||||
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62
    |||||

RESULT 2
Q29032
ID Q29032 PRELIMINARY; PRT; 245 AA.
AC Q29032
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Tumstatin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,
RA Bricksen M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.;
RT "Distinct anti-tumor properties of a type IV collagen domain derived
RT from basement membrane.";
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL; AF258351; AAF72632.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 89.4%; Score 126; DB 4; Length 245;
Best Local Similarity 92.3%; Pred. No. 8.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27
    |||||
DB 70 QRTTTPFLFCNVNDVCFASRNDYS 95
    |||||

RESULT 3
Q29032
ID Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 89.4%; Score 126; DB 4; Length 245;
Best Local Similarity 92.3%; Pred. No. 8.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27
    |||||
DB 70 QRTTTPFLFCNVNDVCFASRNDYS 95
    |||||

RESULT 3
Q29032
ID Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
```

```
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
FT SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 88.7%; Score 125; DB 6; Length 203;
Best Local Similarity 88.5%; Pred. No. 1e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27
    |||||
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62
    |||||
```

```
RESULT 4
Q28682
ID Q28682 PRELIMINARY; PRT; 203 AA.
AC Q28682
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47283; AAA91893.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
FT SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 88.7%; Score 125; DB 6; Length 203;
Best Local Similarity 88.5%; Pred. No. 1e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27
    |||||
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62
    |||||
```

```
RESULT 5
Q28567
ID Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
```

```

GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA31904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4 C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER
FT NON_TER
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 88.7%; Score 125; DB 6; Length 212;
Best Local Similarity 88.5%; Pred. No. 1e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTWPFLFSNVNDVSNFASRNDYS 27
Db ||||| :|||
37 QRTTWPFLFCNINNVCFASRNDYS 62

RESULT 6
Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberhauser I.;
RT "Cloning of the NC1 domains to the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4 C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA31904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4 C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER
FT NON_TER
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 88.7%; Score 125; DB 6; Length 212;
Best Local Similarity 88.5%; Pred. No. 1e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTWPFLFSNVNDVSNFASRNDYS 27
Db ||||| :|||
37 QRTTWPFLFCNINNVCFASRNDYS 62

RESULT 6
Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberhauser I.;
RT "Cloning of the NC1 domains to the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4 C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

Query Match 85.1%; Score 120; DB 11; Length 161;
Best Local Similarity 84.6%; Pred. No. 4.4e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTWPFLFSNVNDVSNFASRNDYS 27
Db ||||| :|||
4 QRTTWPFLFCNINNVCFASRNDYS 29

RESULT 7
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RL MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4 C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA823633D CRC64;

Query Match 85.1%; Score 120; DB 6; Length 210;
Best Local Similarity 84.6%; Pred. No. 5.8e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTWPFLFSNVNDVSNFASRNDYS 27
Db ||||| :|||
47 QRTTWPFLFCNINNVCFASRNDYS 72

RESULT 8
Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=950505957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal

```


RT laminae: Sequence, distribution, association with laminins, and
developmental switches.";
RL J. Cell Biol. 127:879-891(1994).
RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Balb/c;

RA Miner J.H.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z35166; CAA84529.1; -.

DR PIR: I48302; I48302.

DR MGD; MGI:104688; Col4a3.

DR GO: GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR001442; Procollagn4 C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SMO0111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

FT NON_TER 1

SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 85.1%; Score 120; DB 11; Length 246;

Best Local Similarity 84.6%; Pred. No. 6.9e-10;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMEPFLFSNVNDYSNFASRNDYS 27

Db 71 QRTTMEPFLFCNINNVNCFASRNDYS 96

RESULT 9

ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.

AC Q9QZS0;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha 3 collagen IV.

GN COL4A3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=20005934; PubMed=10534397;

RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,

Elder F.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;

RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a

mouse model of alport syndrome.";

RL Genomics 61:113-124(1999).

DR EMBL; AF169387; AAD50449.1; -.

DR PIR; I48302; I48302.

DR MGD; MGI:104688; Col4a3.

DR GO: GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR008161; Clg_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR001442; Procollagn4 C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD000007; Collagen; 21.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SMO0111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

KW Collagen.

SQ SEQUENCE 1669 AA; 161769 MW; 30976B59739A47B2 CRC64;

Query Match

Best Local Similarity 85.1%; Score 120; DB 11; Length 1669;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMEPFLFSNVNDYSNFASRNDYS 27

Db 1494 QRTTMEPFLFCNINNVNCFASRNDYS 1519

RESULT 10

Q63122

ID Q63122 PRELIMINARY; PRT; 230 AA.

AC Q63122;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha-3 type IV collagen (Fragment).

GN COL4A3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

EX MEDLINE=98210005; PubMed=9550634;

RA Ryan J.J., Katbanna I., Mason P.J., Pusey C.D., Turner A.N.;

RT "Sequence analysis of the 'Goodpasture antigen' of mammals.";

RL Nephrol. Dial. Transplant. 13:602-607(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

RA Turner N.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; L47281; AAB72238.2; -.

DR GO: GO:0005581; C:collagen; IEA.

DR GO: GO:000201; F:extracellular matrix structural constituent; IEA.

DR GO: GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollagn4 C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SMO0111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

KW Collagen.

FT NON_TER 1

FT NON_TER 230

SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 82.3%; Score 116; DB 11; Length 230;

Best Local Similarity 84.6%; Pred. No. 2.6e-09;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMEPFLFSNVNDYSNFASRNDYS 27

Db 55 QRTTMEPFLFCNINNVNCFASRNDYS 80

RESULT 11

P70165

ID P70165 PRELIMINARY; PRT; 179 AA.

AC P70165;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)

DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Collagen type IV alpha5 chain (Fragment).

GN COL4A5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129;

RA Oberbauer I.;

RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse

via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and

```
RT alpha5(IV) in differentiated teratocarcinoma cells."
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82218; CAA57698.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1 1
FT NON TER 179 179
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 78.0%; Score 110; DB 11; Length 179;
Best Local Similarity 73.1%; Pred. No. 1.6e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db :|||||:|:|:|||||
32 RRFSTMPFMCNINNVCFASRNDYS 57

RESULT 12
Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: sequence, distribution, association with laminins, and
RT developmental switches."
RL J. Cell Biol. 127:879-891(1994).
DR EMBL; Z35168; CAA84531.1; -.
DR PIR; I48304; I48304.
DR MGD; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1 1
FT NON TER 179 179
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match 78.0%; Score 110; DB 11; Length 253;
Best Local Similarity 73.1%; Pred. No. 2.3e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db :|||||:|:|:|||||
79 RRFSTMPFMCNINNVCFASRNDYS 104

RESULT 13
Q80V57 PRELIMINARY; PRT; 585 AA.
AC Q80V57;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Col4a5 protein.

Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
SEQUENCE FROM N.A.
STRAIN=FVB/N; TISSUE=Breast tumor;
MEDLINE=12477932;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Dhachenko L., Marusina K., Farmer A.S., Rubin G.M., Hong L.,
Stapleton M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Sutterfield Y.S.,
Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
STRAIN=FVB/N; TISSUE=Breast tumor;
Strausberg R.;
Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC043317; AAH43317.1; -.
GO; GO:0005581; C:collagen; IEA.
GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
InterPro; IPR008160; Collagen.
InterPro; IPR001442; Procollagn4_C.
Pfam; PF01413; C4; 2.
Pfam; PF01391; Collagen; 5.
ProDom; PD003923; ProcollagnC4; 2.
SMART; SM00111; C4; 2.
SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match 78.0%; Score 110; DB 11; Length 585;
Best Local Similarity 73.1%; Pred. No. 5.5e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db :|||||:|:|:|||||
411 RRFSTMPFMCNINNVCFASRNDYS 436

RESULT 14
Q8ENS7 PRELIMINARY; PRT; 799 AA.
AC Q8ENS7;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Procollagen (fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Cortex;
MEDLINE=22354683; PubMed=12466851;
The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
"Analysis of the mouse transcriptome based on functional annotation of
```

```

60,770 full-length cDNAs.";
RL Nature 420:583-573(2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGD; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 78.0%; Score 110; DB 11; Length 799;
Best Local Similarity 73.1%; Pred. NO. 7.7e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHPLFSNVNDVNSFASRNDYS 27
DB 625 RRFSTMPFMCNINNVCFASRNDYS 650

RESULT 15
Q9NUE7 PRELIMINARY; PRT; 886 AA.
AC Q9NUE7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 78.0%; Score 110; DB 4; Length 886;
Best Local Similarity 73.1%; Pred. NO. 8.6e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHPLFSNVNDVNSFASRNDYS 27
DB 712 RRFSTMPFMCNINNVCFASRNDYS 737

Search completed: April 5, 2004, 07:03:58
Job time : 16.2131 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 24.3196 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 KQRTTWPFLFSNVNDVSNFASRDYs 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 141 | 100.0 | 27 | ADA20239 | Ada20239 T8-3 pept |
| 2 | 133 | 94.3 | 27 | ADA20241 | Ada20241 P2 peptid |
| 3 | 131 | 92.9 | 27 | ADA20238 | Ada20238 T8 peptid |
| 4 | 126 | 89.4 | 79 | AAU75600 | AAU75600 Human typ |
| 5 | 126 | 89.4 | 79 | ADA20264 | ADA20264 Human tum |
| 6 | 126 | 89.4 | 88 | AAU75608 | AAU75608 Human typ |
| 7 | 126 | 89.4 | 88 | AAU75607 | AAU75607 Human typ |
| 8 | 126 | 89.4 | 88 | ADA20271 | Ada20271 Human tum |
| 9 | 126 | 89.4 | 88 | ADA20272 | Ada20272 Human tum |
| 10 | 126 | 89.4 | 124 | AAU75594 | AAU75594 Human typ |
| 11 | 126 | 89.4 | 124 | ADA20258 | Ada20258 Human tum |
| 12 | 126 | 89.4 | 132 | AAU75597 | AAU75597 Human typ |
| 13 | 126 | 89.4 | 132 | ADA20261 | Ada20261 Human tum |
| 14 | 126 | 89.4 | 191 | AAU75596 | AAU75596 Human typ |
| 15 | 126 | 89.4 | 191 | ADA20260 | Ada20260 Human tum |
| 16 | 126 | 89.4 | 211 | AAU755918 | AAU755918 Human Goo |
| 17 | 126 | 89.4 | 211 | ASG79208 | ABG79208 Human GP |
| 18 | 126 | 89.4 | 218 | AAU79164 | AAU79164 Partial s |
| 19 | 126 | 89.4 | 218 | AAU79164 | AAU79164 Human typ |
| 20 | 126 | 89.4 | 218 | AAU56784 | AAU56784 Human alp |
| 21 | 126 | 89.4 | 218 | AAU09484 | AAU09484 Human alp |
| 22 | 126 | 89.4 | 232 | ADC17697 | ADC17697 Human typ |
| 23 | 126 | 89.4 | 244 | ASG79218 | ABG79218 Human typ |
| 24 | 126 | 89.4 | 244 | ASG79219 | ABG79219 Human Goo |
| 25 | 126 | 89.4 | 244 | ASG79217 | ABG79217 Human typ |

| | | | | | | |
|----|-----|------|------|---|----------|--------------------|
| 26 | 126 | 89.4 | 244 | 5 | AAU75595 | AAU75595 Human typ |
| 27 | 126 | 89.4 | 244 | 6 | ADA20225 | Ada20225 Human typ |
| 28 | 126 | 89.4 | 245 | 3 | AAU67942 | AAU67942 Human typ |
| 29 | 126 | 89.4 | 245 | 5 | AAU75589 | AAU75589 Human typ |
| 30 | 126 | 89.4 | 254 | 5 | AAU75598 | AAU75598 Human typ |
| 31 | 126 | 89.4 | 268 | 2 | AAU31993 | AAU31993 Type IV c |
| 32 | 126 | 89.4 | 268 | 3 | AAU97555 | AAU97555 Human alp |
| 33 | 126 | 89.4 | 1670 | 7 | ADA47063 | Ada47063 Human pro |
| 34 | 125 | 88.7 | 471 | 2 | AAU79163 | AAU79163 Partial s |
| 35 | 125 | 88.7 | 471 | 2 | AAU44171 | AAU44171 Bovine ty |
| 36 | 125 | 88.7 | 471 | 3 | AAU56783 | AAU56783 Bovine al |
| 37 | 125 | 88.7 | 471 | 3 | AAU09483 | AAU09483 Bovine al |
| 38 | 116 | 82.3 | 230 | 7 | ADA47061 | Ada47061 Rat Prote |
| 39 | 110 | 78.0 | 229 | 7 | ADC17699 | ADC17699 Human typ |
| 40 | 110 | 78.0 | 264 | 2 | AAU31995 | AAU31995 Type IV c |
| 41 | 110 | 78.0 | 264 | 3 | AAU97557 | AAU97557 Human alp |
| 42 | 110 | 78.0 | 309 | 3 | AAU54044 | AAU54044 Human pan |
| 43 | 110 | 78.0 | 772 | 2 | AAU23873 | AAU23873 Human alp |
| 44 | 110 | 78.0 | 772 | 2 | AAU09643 | AAU09643 Human typ |
| 45 | 110 | 78.0 | 1685 | 4 | ABG04839 | ABG04839 Novel hum |

ALIGNMENTS

RESULT 1

ADA20239

ID ADA20239 standard; peptide; 27 AA.

AC ADA20239;

DT 20-NOV-2003 (first entry)

DE T8-3 peptide related to human type IV collagen alpha and angiogenesis.

KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytostatic; gene therapy; T8-3 peptide; tumstatin; human;
KW type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

FT Misc-difference 12 /note= "Wild-type Cys substituted by Ser"

FT Misc-difference 18 /note= "Wild-type Cys substituted by Ser"

PN WO2003059257-A2.

XX 24-JUL-2003.

PF 20-DEC-2002; 2002WO-US040938.

PR 21-DEC-2001; 2001US-00032221.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI, 2003-587256/55.

DR New peptide, useful for preparing a composition for inhibiting tumor growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 63; Page 45; 240pp; English.

CC This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the T8-3 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 XX
 SQ Sequence 27 AA;

Query Match 100.0%; Score 141; DB 6; Length 27;
 Best Local Similarity 100.0%; Pred. No. 7.5e-16;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
 DB 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
 |||||

RESULT 2
 ADA20241
 ID ADA20241 standard; peptide; 27 AA.

XX ADA20241;

XX 20-NOV-2003 (first entry)

XX P2 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; P2 peptide; tumstatin; human;
 KW type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.
 OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

FT Misc-difference 12 /note= "Wild-type Cys substituted by Asp"

FT Misc-difference 18 /note= "Wild-type Cys substituted by Asp"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX

PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX
 XX Claim 65; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the P2 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 XX

SQ Sequence 27 AA;

Query Match 94.3%; Score 133; DB 6; Length 27;
 Best Local Similarity 92.6%; Pred. No. 1.6e-14;
 Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
 |||||

DB 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
 |||||

RESULT 3
 ADA20238

ID ADA20238 standard; peptide; 27 AA.

XX ADA20238;

XX 20-NOV-2003 (first entry)

XX T8 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; T8 peptide; tumstatin; human;
 KW type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX

DR WPI; 2003-587256/55.
 XX
 PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX
 PS Claim 62; Page 45; 240pp; English.
 XX
 CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumor growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumor
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the 18 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 XX
 SQ Sequence 27 AA;
 Query Match 92.9%; Score 131; DB 6; Length 27;
 Best Local Similarity 92.6%; Pred. No. 3.3e-14;
 Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 1 KQRTTTPFLFSNVNDVSNFASRNDYS 27
 DB 1 KQRTTTPFLFCNVNDVCFASRNDYS 27
 RESULT 4
 AAU75600
 ID AAU75600 standard; protein; 79 AA.
 XX
 AC AAU75600;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain mutant, Tum-5.
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.
 XX
 OS Homo sapiens.
 XX
 PN WO200151523-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 08-JAN-2001; 2001WO-US000565.
 XX
 PR 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA Kalluri R;
 PI
 XX WPI; 2002-188037/24.
 DR
 XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

PT treating disorders involving angiogenesis.
 XX Example 40; Page; 205pp; English.
 XX
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2, or
 CC alpha3, alpha4, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues
 CC 54-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75589)
 XX
 SQ Sequence 79 AA;
 Query Match 89.4%; Score 126; DB 5; Length 79;
 Best Local Similarity 92.3%; Pred. No. 8.3e-13;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27
 DB 17 QRTTTPFLFCNVNDVCFASRNDYS 42
 RESULT 5
 ADA20264
 ID ADA20264 standard; protein; 79 AA.
 XX
 AC ADA20264;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumstatin deletion protein tum-5 amino acid sequence.
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.
 XX
 OS Homo sapiens.

```

XX PN WO2003059257-A2.
XX PD 24-JUL-2003.
XX PF 20-DEC-2002; 2002WO-US040938.
XX PR 21-DEC-2001; 2001US-00032221.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2003-587256/55.
XX DR N-PSDB; ADA20224.
XX FT New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
PS Claim 94; SEQ ID NO 26; 240pp; English.
XX CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tun-5, an abridged form of the "tumstatin" protein of
CC the invention which was derived from the amino acid sequence of the alpha
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does
CC not appear in the specification but was created by the indexer from
CC information given in the specification.
XX SQ Sequence 79 AA;
    Query Match      89.4%; Score 126; DB 6; Length 79;
    Best Local Similarity 92.3%; Pred. No. 8.3e-13;
    Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRFTTMPFLFSNVNDVSNFASRNDYS 27
DB 16 QRFTTMPFLFCNVNDVCFASRNDYS 41
RESULT 6
AAU75608
ID AAU75608 standard; protein; 88 AA.
XX AC AAU75608;
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 82
/Note= "Wild type Cys substituted with Ala"
WO200151523-A2.
19-JUL-2001.
08-JAN-2001; 2001WO-US000565.
07-JAN-2000; 2000US-00479118.
04-APR-2000; 2000US-00543371.
21-JUL-2000; 2000US-00625191.
(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
Kalluri R;
WPI; 2002-188037/24.
A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
treating disorders involving angiogenesis.
Claim 41; Page 153; 205pp; English.
The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
domain, having one or more of the characteristics selected from: (a) the
ability to bind alphavbeta3 integrin; (b) the ability to inhibit
proliferation of endothelial cells; and (c) the ability to cause
apoptosis of endothelial cells. Also described are the following: (1) use
of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
analogue or allelic variant in the preparation of a medicament for
treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
where the angiogenesis is mediated by one or more endothelial cell
integrins or one or more endothelial cell integrin subunits; or (b) by
promoting or inducing endothelial cell apoptosis in a tissue, where the
endothelial cell apoptosis is mediated by one or more endothelial cell
integrins or one or more endothelial cell integrin subunits; (2) use of
an antibody or peptide that specifically binds the alpha1, alpha2,
alpha3, alpha4, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin
preparation of a medicament for inhibiting angiogenesis or cell
proliferation; (3) use of an inhibitor, such as an antibody, antibody
fragment or peptide of receptor-mediated angiogenesis in the preparation
of a medicament for treating a proliferative disease in a vertebrate,
where the disease is characterised by angiogenesis that is mediated by
receptors to Arresten, Canstatin or Tumstatin and where the receptors
inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
the presence of a medicament for promoting angiogenesis in a tissue; and
(5) use of integrins in the preparation of a medicament for promoting or
inducing angiogenesis or cell proliferation in a tissue. The fragments
Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
or allelic variants are useful in the preparation of a medicament for
treating a disorder involving inhibiting angiogenesis in a tissue, where
the angiogenesis is mediated by one or more endothelial cell integrins or
one or more endothelial cell integrin subunits; or by promoting or
inducing endothelial cell apoptosis in a tissue, where the endothelial
cell apoptosis is mediated by one or more endothelial cell integrins or
one or more endothelial cell integrin subunits. The medicament is useful
in inhibiting tumour growth and for the regression of an established
tumour. The present sequence represents the amino acid sequence of human
type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
consists of residues 5-126 of Tumstatin
Sequence 88 AA;
Query Match      89.4%; Score 126; DB 5; Length 88;
Best Local Similarity 92.3%; Pred. No. 9.5e-13;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRFTTMPFLFSNVNDVSNFASRNDYS 27
DB 26 QRFTTMPFLFCNVNDVCFASRNDYS 51

```

RESULT 7
AAU75607
XX AAU75607 standard; protein; 88 AA.
XX AC
XX AAU75607;
XX AC
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.
XX XX
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX Tumstatin; angiogenesis; tumour; mutant.
XX XX
XX Homo sapiens.
XX XX
XX WO200151523-A2.
XX XX
XX PD 19-JUL-2001.
XX XX
XX PF 08-JAN-2001; 2001WO-US000565.
XX XX
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX XX
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX XX
XX WPI; 2002-188037/24.
XX XX
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX PT treating disorders involving angiogenesis.
XX PT
XX PS Claim 32; Page 152; 205pp; English.
XX XX
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; or (b) by
XX promoting or inducing endothelial cell apoptosis in a tissue, where the
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in the preparation
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX or allelic variants are useful in the preparation of a medicament for
XX treating a disorder involving inhibiting angiogenesis in a tissue, where
XX the angiogenesis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits; or by promoting or
XX inducing endothelial cell apoptosis in a tissue, where the endothelial
XX cell apoptosis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits. The medicament is useful
XX in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists
CC of residues 45-132 of Tumstatin
XX XX
XX SQ Sequence 88 AA;
XX
XX Query Match 89,4%; Score 126; DB 5; Length 88;
XX Best Local Similarity 92,3%; Pred. No. 9,5e-13;
XX Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 QRFTTMBPLFSNVNDVGNFASRNDYS 27
XX ||||| ||||| ||||| ||||| |||||
XX Db 26 QRFTTMBPLFCNVNDVGNFASRNDYS 51
XX
XX RESULT 8
XX ADA20271
XX ID ADA20271 standard; protein; 88 AA.
XX AC
XX AC ADA20271;
XX XX
XX DT 20-NOV-2003 (first entry)
XX XX
XX DE Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human;
XX tumstatin 45-132.
XX XX
XX OS Homo sapiens.
XX XX
XX FN WO20003059257-A2.
XX XX
XX PD 24-JUL-2003.
XX XX
XX PF 20-DEC-2002; 2002WO-US040938.
XX PR 21-DEC-2001; 2001US-00032221.
XX XX
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2003-587256/55.
XX DR N-PSDB; ADA20224.
XX XX
XX PT New peptide, useful for preparing a composition for inhibiting tumour
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX PS
XX Claim 94; SEQ ID NO 33; 240pp; English.
XX XX
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"
XX protein of the invention which was derived from the amino acid sequence
XX of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq

CC ID33) does not appear in the specification but was created by the indexer
 CC from information given in the specification.

XX SQ Sequence 88 AA;

Query Match 89.4%; Score 126; DB 6; Length 88;
 Best Local Similarity 92.3%; Pred. No. 9.5e-13;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTHMPFLPSNVDSNFASRNDYS 27
 DB 25 QRFTHMPFLPSNVDSNFASRNDYS 50

RESULT 9

ID ADA20272 standard; protein; 88 AA.

XX AC ADA20272;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX KW metastasis; basement membrane organisation; type IV collagen network;

XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX KW cytotactic; gene therapy; alpha 3 chain; tumstatin; human;

XX KW tumstatin 5-125-C-A; mutant; murein.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 81

XX FT /note= "Wild-type Cys substituted by Ala at position 125

XX FT of full-length tumstatin"

XX PN WO2003059257-A2.

XX XX 24-JUL-2003.

XX XX 20-DEC-2002; 2002WO-US040938.

XX PF 21-DEC-2001; 2001US-00032221.

XX PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PA Kalluri R;

XX PI WPI; 2002-188037/24.

XX DR WPI; 2003-587256/55.

XX XX New peptide, useful for preparing a composition for inhibiting tumor

XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 34; 240pp; English.

XX XX This invention relates to novel isolated proteins and their fragments

XX CC with anti-angiogenic properties. The invention also relates to the DNA

XX CC sequences which encode the novel proteins. A wide variety of diseases are

XX CC the result of undesirable angiogenesis. The formation of new capillaries

XX CC from pre-existing vessels is essential for tumour growth and metastasis.

XX CC Basement membrane organisation is dependent on the assembly of a type IV

XX CC collagen network which may occur through the C-terminal globular non-

XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

XX CC forms are ubiquitously exhibited in human basement membranes. In the

XX CC present invention, cell surface receptors (in particular integrins) which

XX CC specifically bind anti-angiogenic proteins and peptides (in particular

XX CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

XX CC collagen) are disclosed. The proteins of the invention may inhibit tumour

XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in

XX CC endothelial cells and thus may exhibit cytotactic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
 CC the "tumstatin" protein of the invention which was derived from the amino
 CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This
 CC sequence (Seq ID33) does not appear in the specification but was created
 CC by the indexer from information given in the specification.

XX SQ Sequence 88 AA;

Query Match 89.4%; Score 126; DB 6; Length 88;
 Best Local Similarity 92.3%; Pred. No. 9.5e-13;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTHMPFLPSNVDSNFASRNDYS 27
 DB 25 QRFTHMPFLPSNVDSNFASRNDYS 50

RESULT 10

AAU75594

ID AAU75594 standard; protein; 124 AA.

XX AC AAU75594;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.

XX KW Human; type IV collagen alpha 3 chain; cytotactic; antiangiogenic;

XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;

XX KW Tumstatin; angiogenesis; tumour; murein; mutant.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US0000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX PI WPI; 2002-188037/24.

XX DR WPI; 2002-188037/24.

XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

XX PT treating disorders involving angiogenesis.

XX PS Example 33; Page; 205pp; English.

XX XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

XX CC domain, having one or more of the characteristics selected from: (a) the

XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

XX CC proliferation of endothelial cells; and (c) the ability to cause

XX CC apoptosis of endothelial cells. Also described are the following: (1) use

XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

XX CC analogue or allelic variant in the preparation of a medicament for

XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

XX CC where the angiogenesis is mediated by one or more endothelial cell

XX CC integrins or one or more endothelial cell integrin subunits; or (b) by

XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the

XX CC endothelial cell apoptosis is mediated by one or more endothelial cell

XX CC integrins or one or more endothelial cell integrin subunits; (2) use of

XX CC an antibody or peptide that specifically binds the alpha1, alpha2,

XX CC alpha3, alpha4, alpha6, alphav, betav or beta3 subunit of integrin in the

XX CC preparation of a medicament for inhibiting angiogenesis or cell

XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody

CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of
 CC residues 2-125 of tumstatin. Note: The present sequence is not shown in
 CC the specification but is derived from the wild type human Tumstatin
 CC sequence given in figure 18A (see AAU75589)

XX SQ Sequence 124 AA;

Query Match 89.4%; Score 126; DB 5; Length 124;
 Best Local Similarity 92.3%; Pred. No. 1.5e-12;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFITMPFLFSNVNDVSNFASRNDYS 27
 |||||
 Db 69 QRFITMPFLFCNVNDVCFASRNDYS 94

RESULT 11

ADA20258
 ID ADA20258 standard; protein; 124 AA.

XX AC ADA20258;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 XX KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytosstatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 20; 240pp; English.

CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"
 CC protein of the invention which was derived from the amino acid sequence
 CC of the alpha 3 chain of human type IV collagen. Note: This sequence (seq
 CC ID20) does not appear in the specification but was created by the indexer
 CC from information given in the specification.

XX SQ Sequence 124 AA;

Query Match 89.4%; Score 126; DB 6; Length 124;
 Best Local Similarity 92.3%; Pred. No. 1.5e-12;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFITMPFLFSNVNDVSNFASRNDYS 27
 |||||
 Db 69 QRFITMPFLFCNVNDVCFASRNDYS 94

RESULT 12

AAU75597

ID AAU75597 standard; protein; 132 AA.

XX AC AAU75597;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tum-2.

XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2002-188037/24.

XX PT A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
 PT treating disorders involving angiogenesis.

XX PS Claim 31; Page 152; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI
 CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, betav or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75589)

XX Sequence 132 AA;

Query Match 89.4%; Score 126; DB 5; Length 132;
 Best Local Similarity 92.3%; Pred. No. 1.6e-12;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTTMPFLFSNVNDVSNFASRNDYS 27
 |||||
 Db 70 QRFTTMPFLFCNVNDVCFNFSRNDYS 95
 |||||

RESULT 13
 ADA20261
 ID ADA20261 standard; protein; 132 AA.

XX ADA20261;

DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; Ncl; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytotatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

PF 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCl) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCl domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.

XX Sequence 132 AA;

Query Match 89.4%; Score 126; DB 6; Length 132;
 Best Local Similarity 92.3%; Pred. No. 1.6e-12;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTTMPFLFSNVNDVSNFASRNDYS 27
 |||||
 Db 69 QRFTTMPFLFCNVNDVCFNFSRNDYS 94
 |||||

RESULT 14

AAU75586
 ID AAU75586 standard; protein; 191 AA.

XX AAU75586;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NCl domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

```

PR 21-JUL-2000; 2000US-00625191.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2002-188037/24.
XX
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX
XX Example 32; Page; 205pp; English.
XX
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; or (b) by
XX promoting or inducing endothelial cell apoptosis in a tissue, where the
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in the preparation
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX or allelic variants are useful in the preparation of a medicament for
XX treating a disorder involving inhibiting angiogenesis in a tissue, where
XX the angiogenesis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits; or by promoting or
XX inducing endothelial cell apoptosis in a tissue, where the endothelial
XX cell apoptosis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits. The medicament is useful
XX in inhibiting tumour growth and for the regression of an established
XX tumour. The present sequence represents the amino acid sequence of human
XX type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of
XX residues 54-244 of Tumstatin. Note: The present sequence is not shown in
XX the specification but is derived from the wild type human Tumstatin
XX sequence given in figure 18A (see AAU75589)
XX
XX Sequence 191 AA;
XX
XX Query Match 89.4%; Score 126; DB 5; Length 191;
XX Best Local Similarity 92.3%; Pred. No. 2.5e-12;
XX Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
XX 17 QRFTTTPFLFSNVNDVSNFASRNDYS 42
XX
XX RESULT 15
XX ADA20260
XX ADA20260 standard; protein; 191 AA.
XX
XX ADA20260;
XX
XX 20-NOV-2003 (first entry)
XX

```

```

DE Human tumstatin deletion protein tum-1 amino acid sequence.
XX
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;
XX tumstatin N53.
XX
XX Homo sapiens.
XX
XX WO2003059257-A2.
XX
XX 24-JUL-2003.
XX
XX 20-DEC-2002; 2002WO-US040938.
XX
XX 21-DEC-2001; 2001US-00032221.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2003-587256/55.
XX N-PSDB; ADA20224.
XX
XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
XX Claim 94; SEQ ID NO 22; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX "tumstatin" protein of the invention which was derived from the amino
XX acid sequence of the alpha 3 chain of human type IV collagen. Note: This
XX sequence (Seq ID22) does not appear in the specification but was created
XX by the indexer from information given in the specification.
XX
XX Sequence 191 AA;
XX
XX Query Match 89.4%; Score 126; DB 6; Length 191;
XX Best Local Similarity 92.3%; Pred. No. 2.5e-12;
XX Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
XX 16 QRFTTTPFLFSNVNDVSNFASRNDYS 41
XX
XX Search completed: April 5, 2004, 06:58:32
XX Job time : 24.3196 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 16.9322 Seconds
(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-40
Perfect score: 141
Sequence: 1 KQRTTMTFLFNSVNDVSNFASRNDYS 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/CT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/CT06_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 141 | 100.0 | 27 | 14 | US-10-032-221B-40 |
| 2 | 133 | 94.3 | 27 | 14 | Sequence 42, Appl |
| 3 | 131 | 92.9 | 27 | 14 | Sequence 39, Appl |
| 4 | 126 | 89.4 | 79 | 14 | Sequence 26, Appl |
| 5 | 126 | 89.4 | 88 | 14 | Sequence 33, Appl |
| 6 | 126 | 89.4 | 124 | 14 | Sequence 34, Appl |
| 7 | 126 | 89.4 | 132 | 14 | Sequence 20, Appl |
| 8 | 126 | 89.4 | 132 | 14 | Sequence 23, Appl |
| 9 | 126 | 89.4 | 131 | 14 | Sequence 22, Appl |
| 10 | 126 | 89.4 | 211 | 14 | Sequence 46, Appl |
| 11 | 126 | 89.4 | 211 | 14 | Sequence 45, Appl |
| 12 | 126 | 89.4 | 232 | 14 | Sequence 304, Appl |
| 13 | 126 | 89.4 | 244 | 14 | Sequence 10, Appl |
| 14 | 110 | 78.0 | 229 | 14 | Sequence 306, Appl |
| 15 | 110 | 78.0 | 309 | 9 | US-09-925-297-496 |

| | | | | | |
|----|-----|------|------|----|---------------------|
| 16 | 109 | 77.3 | 229 | 14 | US-10-206-699-302 |
| 17 | 109 | 77.3 | 229 | 14 | US-10-032-221B-2 |
| 18 | 109 | 77.3 | 408 | 9 | US-09-925-304-507 |
| 19 | 109 | 77.3 | 1669 | 15 | US-10-372-683-8 |
| 20 | 105 | 74.5 | 22 | 14 | US-10-206-699-266 |
| 21 | 105 | 74.5 | 25 | 14 | US-10-032-221B-37 |
| 22 | 97 | 68.8 | 25 | 14 | US-10-032-221B-38 |
| 23 | 95 | 67.4 | 22 | 14 | US-10-206-699-265 |
| 24 | 95 | 67.4 | 1759 | 15 | US-10-369-493-7032 |
| 25 | 93 | 66.0 | 22 | 14 | US-10-206-699-267 |
| 26 | 90 | 63.8 | 20 | 14 | US-10-206-699-289 |
| 27 | 90 | 63.8 | 20 | 14 | US-10-032-221B-29 |
| 28 | 87 | 61.7 | 46 | 9 | US-09-864-761-48095 |
| 29 | 87 | 61.7 | 1744 | 15 | US-10-369-493-5832 |
| 30 | 85 | 60.3 | 142 | 9 | US-09-864-761-38021 |
| 31 | 85 | 60.3 | 228 | 14 | US-10-206-699-307 |
| 32 | 84 | 59.6 | 18 | 14 | US-10-206-699-254 |
| 33 | 84 | 59.6 | 18 | 14 | US-10-206-699-260 |
| 34 | 82 | 58.2 | 227 | 14 | US-10-206-699-303 |
| 35 | 82 | 58.2 | 227 | 14 | US-10-032-221B-6 |
| 36 | 82 | 58.2 | 430 | 9 | US-09-925-302-518 |
| 37 | 82 | 58.2 | 459 | 15 | US-10-331-496A-27 |
| 38 | 82 | 58.2 | 459 | 15 | US-10-372-683-30 |
| 39 | 82 | 58.2 | 1712 | 10 | US-09-961-403-9 |
| 40 | 78 | 55.3 | 18 | 14 | US-10-206-699-259 |
| 41 | 78 | 55.3 | 22 | 14 | US-10-206-699-270 |
| 42 | 76 | 53.9 | 18 | 14 | US-10-206-699-261 |
| 43 | 75 | 53.2 | 22 | 14 | US-10-206-699-268 |
| 44 | 74 | 52.5 | 18 | 14 | US-10-206-699-253 |
| 45 | 74 | 52.5 | 20 | 14 | US-10-206-699-290 |

ALIGNMENTS

RESULT 1

US-10-032-221B-40
; Sequence 40, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T8-3 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length run
; OTHER INFORMATION: statin molecule, and serine has been substituted for the cysteine
; OTHER INFORMATION: residues at positions 79 and 85)
US-10-032-221B-40

Query Match 100.0%; Score 141; DB 14; Length 27;

```
Best Local Similarity 100.0%; Pred. No. 4.2e-15; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0;

Qy 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
      |||||
Db 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
      |||||

RESULT 2
US-10-032-221B-42
; Sequence 42, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: P2 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted
; OTHER INFORMATION: ed for the leucine residue at position 68 of the full-length Tumst
; OTHER INFORMATION: tatin molecule, and aspartic acid has been substituted for the cy
; OTHER INFORMATION: steine residues at positions 79 and 85)
US-10-032-221B-42

Query Match 94.3%; Score 133; DB 14; Length 27;
Best Local Similarity 92.6%; Pred. No. 7.6e-14;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
      |||||
Db 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
      |||||

RESULT 3
US-10-032-221B-39
; Sequence 39, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 39
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 26
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26

Query Match 89.4%; Score 126; DB 14; Length 79;
Best Local Similarity 92.3%; Pred. No. 3.3e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTTPFLFSNVNDSNFASRNDYS 27
      |||||
Db 16 QRTTTPFLFSNVNDSNFASRNDYS 41
      |||||

RESULT 5
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1
```

```
PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 39
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T8 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted
; OTHER INFORMATION: d for the leucine residue at position 68 of the full-length Tumst
; OTHER INFORMATION: atin molecule)
US-10-032-221B-39

Query Match 92.9%; Score 131; DB 14; Length 27;
Best Local Similarity 92.6%; Pred. No. 1.6e-13;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
      |||||
Db 1 KQRTTTPFLFSNVNDSNFASRNDYS 27
      |||||

RESULT 4
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 26
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26

Query Match 89.4%; Score 126; DB 14; Length 79;
Best Local Similarity 92.3%; Pred. No. 3.3e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTTPFLFSNVNDSNFASRNDYS 27
      |||||
Db 16 QRTTTPFLFSNVNDSNFASRNDYS 41
      |||||

RESULT 5
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)
US-10-032-221B-33

Query Match      89.4%; Score 126; DB 14; Length 88;
Best Local Similarity 92.3%; Pred. No. 3.7e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFITMPFLFSNVNDVSNFASRNDYS 27
      |||||
Db      25 QRFITMPFLFCNVNDVCNFASTRNDYS 50

RESULT 6
US-10-032-221B-34
; Sequence 34, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine
; OTHER INFORMATION: has been substituted for the cysteine residue at position 125 of
; OTHER INFORMATION: the full-length Tumstatin molecule)

```

US-10-032-221B-34

```

Query Match      89.4%; Score 126; DB 14; Length 88;
Best Local Similarity 92.3%; Pred. No. 3.7e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFITMPFLFSNVNDVSNFASRNDYS 27
      |||||
Db      25 QRFITMPFLFCNVNDVCNFASTRNDYS 50

```

RESULT 7

```

US-10-032-221B-20
; Sequence 20, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

```

```

Query Match      89.4%; Score 126; DB 14; Length 124;
Best Local Similarity 92.3%; Pred. No. 5.5e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 QRFITMPFLFSNVNDVSNFASRNDYS 27
      |||||
Db      69 QRFITMPFLFCNVNDVCNFASTRNDYS 94

```

RESULT 8

```

US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224

```

```

; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match      89.4%; Score 126; DB 14; Length 132;
Best Local Similarity 92.3%; Pred. No. 6e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRETTMPFLFSNVNDVSNFASRNDYS 27
Db      69 QRETTMPFLFCNVNDVSNFASRNDYS 94

RESULT 9
US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match      89.4%; Score 126; DB 14; Length 191;
Best Local Similarity 92.3%; Pred. No. 9.1e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRETTMPFLFSNVNDVSNFASRNDYS 27
Db      16 QRETTMPFLFCNVNDVSNFASRNDYS 41

RESULT 10
US-10-032-221B-21
; Sequence 46, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27

```

```

; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match      89.4%; Score 126; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRETTMPFLFSNVNDVSNFASRNDYS 27
Db      69 QRETTMPFLFCNVNDVSNFASRNDYS 94

RESULT 11
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054489A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match      89.4%; Score 126; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRETTMPFLFSNVNDVSNFASRNDYS 27
Db      69 QRETTMPFLFCNVNDVSNFASRNDYS 94

RESULT 12
US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27

```


; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 304
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 3 chain
US-10-206-699-304

Query Match 89.4%; Score 126; DB 14; Length 244;
Best Local Similarity 92.3%; Pred. No. 1.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27
|||||
DB 57 QRFTHMPFLFCNVDSNFASRNDYS 82
|||||

RESULT 13
US-10-032-221B-10
; Sequence 10, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram

; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-10

Query Match 89.4%; Score 126; DB 14; Length 244;
Best Local Similarity 92.3%; Pred. No. 1.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27
|||||
DB 69 QRFTHMPFLFCNVDSNFASRNDYS 94
|||||

RESULT 14
US-10-206-699-306
; Sequence 306, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 306
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 5 chain
US-10-206-699-306

Query Match 78.0%; Score 110; DB 14; Length 229;
Best Local Similarity 73.1%; Pred. No. 3.7e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27
|||||
DB 55 RRFSTMPFMCNINNVCFASRNDYS 80
|||||

RESULT 15
US-09-925-297-496
; Sequence 496, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 496
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (247)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-496

Query Match 78.0%; Score 110; DB 9; Length 309;
Best Local Similarity 73.1%; Pred. No. 5.2e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27
|||||
DB 135 RRFSTMPFMCNINNVCFASRNDYS 160
|||||

Search completed: April 5, 2004, 07:36:07
Job time : 17.9322 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 6.3414 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221b-40

Perfect score: 141

Sequence: 1 QKFTTTPFLFNVNDVNFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCUTUS COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 126 | 89.4 | 211 | 4 | US-09-512-563C-46 |
| 2 | 126 | 89.4 | 218 | 2 | US-08-399-889-25 |
| 3 | 126 | 89.4 | 218 | 3 | US-09-167-364-25 |
| 4 | 126 | 89.4 | 218 | 3 | US-09-439-897-4 |
| 5 | 126 | 89.4 | 268 | 4 | US-09-589-927-6 |
| 6 | 126 | 89.4 | 268 | 4 | US-09-277-665-6 |
| 7 | 126 | 89.4 | 268 | 4 | US-09-589-987-6 |
| 8 | 125 | 88.7 | 471 | 2 | US-08-399-889-24 |
| 9 | 125 | 88.7 | 471 | 3 | US-09-167-364-24 |
| 10 | 125 | 88.7 | 471 | 3 | US-09-439-897-2 |
| 11 | 110 | 78.0 | 264 | 4 | US-09-589-927-10 |
| 12 | 110 | 78.0 | 264 | 4 | US-09-277-665-10 |
| 13 | 110 | 78.0 | 264 | 4 | US-09-589-987-10 |
| 14 | 109 | 77.3 | 260 | 4 | US-09-589-927-2 |
| 15 | 109 | 77.3 | 260 | 4 | US-09-277-665-2 |
| 16 | 109 | 77.3 | 260 | 4 | US-09-589-987-2 |
| 17 | 85 | 60.3 | 260 | 4 | US-09-589-927-12 |
| 18 | 85 | 60.3 | 260 | 4 | US-09-277-665-12 |
| 19 | 85 | 60.3 | 260 | 4 | US-09-589-987-12 |
| 20 | 82 | 58.2 | 258 | 4 | US-09-589-927-4 |
| 21 | 82 | 58.2 | 258 | 4 | US-09-277-665-4 |
| 22 | 82 | 58.2 | 258 | 4 | US-09-589-987-4 |
| 23 | 68 | 48.2 | 260 | 4 | US-09-589-927-8 |
| 24 | 68 | 48.2 | 260 | 4 | US-09-277-665-8 |
| 25 | 68 | 48.2 | 260 | 4 | US-09-589-987-8 |
| 26 | 48 | 34.0 | 1694 | 1 | US-08-494-168-2 |
| 27 | 46 | 32.6 | 107 | 3 | US-09-102-528-23 |

| | | | | | | |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 46 | 32.6 | 107 | 3 | US-09-102-528-27 | Sequence 27, Appl |
| 29 | 46 | 32.6 | 286 | 4 | US-08-252-991A-20515 | Sequence 20515, A |
| 30 | 46 | 32.6 | 587 | 3 | US-09-102-528-30 | Sequence 30, Appl |
| 31 | 46 | 32.6 | 736 | 3 | US-09-102-528-29 | Sequence 29, Appl |
| 32 | 45 | 31.9 | 1117 | 4 | PCT-US95-13813-9 | Sequence 9, Appl |
| 33 | 44 | 31.2 | 167 | 5 | US-09-339-159B-4 | Sequence 4, Appl |
| 34 | 44 | 31.2 | 476 | 4 | US-09-198-956-10 | Sequence 10, Appl |
| 35 | 44 | 31.2 | 493 | 3 | US-09-198-956A-12 | Sequence 12, Appl |
| 36 | 44 | 31.2 | 493 | 4 | US-09-694-531-12 | Sequence 12, Appl |
| 37 | 44 | 31.2 | 493 | 4 | US-09-670-141-10 | Sequence 10, Appl |
| 38 | 44 | 31.2 | 493 | 4 | US-10-072-152-12 | Sequence 12, Appl |
| 39 | 44 | 31.2 | 793 | 2 | US-08-468-558-5 | Sequence 5, Appl |
| 40 | 44 | 31.2 | 793 | 3 | US-08-676-444-5 | Sequence 11, Appl |
| 41 | 44 | 31.2 | 793 | 3 | US-08-424-641B-11 | Sequence 11, Appl |
| 42 | 43.5 | 30.9 | 288 | 2 | US-08-820-980-11 | Sequence 11, Appl |
| 43 | 43.5 | 30.9 | 288 | 2 | US-08-826-439-11 | Sequence 11, Appl |
| 44 | 43.5 | 30.9 | 288 | 2 | US-09-489-039A-13105 | Sequence 13105, A |
| 45 | 43 | 30.5 | 192 | 4 | | |

ALIGNMENTS

RESULT 1
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 89.4%; Score 126; DB 4; Length 211;
Best Local Similarity 92.3%; Pred. No. 3.3e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QKFTTTPFLFNVNDVNFASRNDYS 27
DB 69 QKFTTTPFLFNVNDVNFASRNDYS 94

RESULT 2
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT

```
; ORGANISM: Human
US-08-399-889-25

Query Match      89.4%; Score 126; DB 2; Length 218;
Best Local Similarity 92.3%; Pred. No. 3.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match      89.4%; Score 126; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. No. 3.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      89.4%; Score 126; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. No. 3.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match      89.4%; Score 126; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 4.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match      89.4%; Score 126; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 4.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match      89.4%; Score 126; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 4.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118
```

```
Best Local Similarity 92.3%; Pred. No. 4.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118

RESULT 8
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reiders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match 88.7%; Score 125; DB 2; Length 471;
Best Local Similarity 88.5%; Pred. No. 1.2e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 296 QRTTMPFLFCNVNDVCFASRNDYS 321

RESULT 9
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reiders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match 88.7%; Score 125; DB 3; Length 471;
Best Local Similarity 88.5%; Pred. No. 1.2e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 296 QRTTMPFLFCNVNDVCFASRNDYS 321

RESULT 10
US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 88.7%; Score 125; DB 3; Length 471;
Best Local Similarity 88.5%; Pred. No. 1.2e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 296 QRTTMPFLFCNVNDVCFASRNDYS 321

RESULT 11
US-09-589-927-10
; Sequence 10, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-10

Query Match 78.0%; Score 110; DB 4; Length 264;
Best Local Similarity 73.1%; Pred. No. 1.4e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
Db 90 RRFSTMPFECNINNVCFASRNDYS 115

RESULT 12
US-09-277-665-10
; Sequence 10, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-10

Query Match 78.0%; Score 110; DB 4; Length 264;
Best Local Similarity 73.1%; Pred. No. 1.4e-09;
```

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27
:|||||:|:|:|||||
Db 90 RRFSTMPFMCNINNVCFASRNDYS 115

RESULT 13

US-09-589-987-10
; Sequence 10, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-10

Query Match 78.0%; Score 110; DB 4; Length 264;
Best Local Similarity 73.1%; Pred. No. 1.4e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27
:|||||:|:|:|||||
Db 90 RRFSTMPFMCNINNVCFASRNDYS 115

RESULT 14

US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 77.3%; Score 109; DB 4; Length 260;
Best Local Similarity 73.1%; Pred. No. 2e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27
:|||||:|:|:|||||
Db 86 RRFSTMPFMCNINNVCFASRNDYS 111

RESULT 15

US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2

Query Match 77.3%; Score 109; DB 4; Length 260;
Best Local Similarity 73.1%; Pred. No. 2e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27
:|||||:|:|:|||||
Db 86 RRFSTMPFMCNINNVCFASRNDYS 111

Search completed: April 5, 2004, 07:07:25
Job time : 6.3414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 3.91041 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 LFCNVNVCNCFASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pirl:*

2: Pirl:*

3: Pirl:*

4: Pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|---------------------|
| 1 | 93 | 84.5 | 220 | B49736 | collagen alpha 3(I) |
| 2 | 93 | 84.5 | 1670 | CGH3B | collagen alpha 3(I) |
| 3 | 92 | 83.6 | 161 | S49488 | collagen alpha 3(I) |
| 4 | 92 | 83.6 | 246 | I48302 | collagen alpha 3(I) |
| 5 | 92 | 83.6 | 258 | B61228 | collagen alpha 3(I) |
| 6 | 92 | 83.6 | 471 | A39024 | collagen alpha 3(I) |
| 7 | 92 | 83.6 | 1669 | CGH4B | collagen alpha 1(I) |
| 8 | 92 | 83.6 | 1669 | CGH34B | collagen alpha 1(I) |
| 9 | 90 | 81.8 | 253 | I48304 | collagen alpha 5(I) |
| 10 | 90 | 81.8 | 754 | A55267 | collagen alpha 5(I) |
| 11 | 90 | 81.8 | 1691 | S22917 | collagen alpha 5(I) |
| 12 | 80 | 72.7 | 1752 | A45407 | collagen alpha 3(I) |
| 13 | 76 | 69.1 | 1747 | A54121 | collagen alpha-4 c |
| 14 | 75 | 69.1 | 1763 | S16366 | collagen alpha 2(I) |
| 15 | 75 | 68.2 | 1744 | S40391 | collagen alpha 1(I) |
| 16 | 71 | 64.5 | 1758 | T29350 | hypothetical prote |
| 17 | 71 | 64.5 | 1759 | T29351 | collagen alpha 2(I) |
| 18 | 67 | 60.9 | 261 | A34476 | collagen alpha 2(I) |
| 19 | 64 | 58.2 | 1691 | CGH6B | collagen alpha 6(I) |
| 20 | 59 | 53.6 | 312 | I48303 | collagen alpha 4(I) |
| 21 | 59 | 53.6 | 623 | A45137 | collagen alpha 4(I) |
| 22 | 59 | 53.6 | 1690 | CGH2B | collagen alpha 4(I) |
| 23 | 59 | 53.6 | 1755 | A31893 | collagen alpha 1(I) |
| 24 | 58 | 52.7 | 453 | S18804 | collagen alpha 4(I) |
| 25 | 58 | 52.7 | 775 | A61228 | collagen alpha 2(I) |
| 26 | 58 | 52.7 | 1707 | A33526 | collagen alpha 2(I) |
| 27 | 58 | 52.7 | 1712 | CGH2B | collagen alpha 2(I) |
| 28 | 53 | 48.2 | 79 | C43928 | probable collagen |
| 29 | 43 | 44.5 | 346 | T46914 | hypothetical prote |

30 49 44.5 438 2 E84579
31 49 44.5 1761 2 T13990
32 47 42.7 1477 2 T18534
33 46.5 42.3 419 2 S41607
34 46 41.8 257 2 H75419
35 46 41.8 338 2 T05036
36 46 41.8 760 2 T41644
37 45 40.9 840 2 T38528
38 43 39.1 114 2 T18089
39 43 39.1 1270 2 T51227
40 43 39.1 2946 2 T00867
41 42.5 38.6 553 2 B72863
42 42.5 38.6 578 2 S50446
43 42.5 38.6 869 2 F97126
44 42.5 38.6 1270 2 T26720
45 42 38.2 75 2 JC6048

ALIGNMENTS

RESULT 1

B49736
collagen alpha 3(IV) chain, medium splice form - human (fragment)

N:Contains: collagen alpha 3(IV) chain, splice form GP-V

C:Species: Homo sapiens (man)

C>Date: 03-May-1994 #sequence revision 12-Nov-1999 #text_change 17-Mar-2000

C:Accession: B49736; D49736; S69111

R:Feng, L.; Xia, Y.; Wilson, C.B.

J. Biol. Chem. 269, 2342-2348, 1994

A:Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.

A:Reference number: A49736; MUID:94124597; PMID:8294492

A:Accession: B49736

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 169-220 <FEN1>

A:Accession: D49736

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 22-220 <FEN2>

A:Cross-references: GB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107

A>Note: this is the conceptual translation of the nucleic acid submitted to GenBank

R:Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; W

Eur. J. Biochem. 229, 754-760, 1995

A:Title: Characterization and expression of multiple alternatively spliced transcripts

into antigen and one of its alternative forms.

A:Reference number: S69111; MUID:95278230; PMID:7758473

A:Accession: S69111

A:Molecule type: mRNA

A:Residues: 1-45,169-204,'L',206-220 <PEN>

C:Comment: For the complete sequence of the long splice form, see PIR:CGH3B.

C:Genetics:

A:Gene: GDB:COL4A3

A:Cross-references: GDB:128351; OMIM:120070

A:Map position: 2q36-2q37

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrac

F:1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pre

F:1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #statu

F:22-220/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

F:134-134/Domain: collagen IV carboxyl-terminal repeat <Ctrl>

Query Match 84.5%; Score 93; DB 2; Length 220;

Best Local Similarity 94.4%; Pred. No. 2,7e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19

DB 86 LFCNVNVCNCFASRNDYS 103

RESULT 2

CGH3B

collagen alpha 3(IV) chain precursor, long splice form - human
 N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
 C/Species: Homo sapiens (man)
 C/Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 22-Jun-1999
 C/Accession: A54763; A43928; A44043; A45971; A39786
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.
 J. Biol. Chem. 269, 23013-23017, 1994
 A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
 A/Reference number: A54763; MUID:94364994; PMID:8083201
 A/Accession: A54763
 A/Molecule type: mRNA
 A/Residues: 1-1670 <NAR>
 A/Cross-references: GB:X80031; NID:G577563; PID:G577564
 A/Experimental source: Kidney
 R/Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
 J. Clin. Invest. 89, 592-601, 1992
 A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha
 A/Reference number: A43928; MUID:92147878; PMID:1737849
 A/Accession: A43928
 A/Molecule type: mRNA
 A/Residues: 1331-1524, 1526-1670 <TUR>
 A/Cross-references: GB:M81379
 A/Experimental source: Kidney
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 267, 19780-19784, 1992
 A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
 cation.
 A/Reference number: A44043; MUID:93015826; PMID:1400291
 A/Accession: A44043
 A/Molecule type: DNA; mRNA
 A/Residues: 1386-1670 <QUI>
 A/Cross-references: GB:X92993; NID:G177895; PID:AAA21610.1; PID:G177896
 A/Note: sequence extracted from NCBI backbone (NCBI:P115597)
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 269, 17358, 1994
 A/Reference number: A44738; MUID:94274734; PMID:8006044
 A/Contents: annotation; erratum; correction to intronic sequence in A44043
 R/Bernal, D.; Quinones, S.; Saus, J.
 J. Biol. Chem. 268, 12090-12094, 1993
 A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A/Reference number: A45971; MUID:93280184; PMID:8505332
 A/Accession: A45971
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1427-1444 <BER>
 A/Note: sequence extracted from NCBI backbone (NCBI:P133363); sequence incorrectly identified
 R/Morrison, K.E.; Mariyama, M.; Yang-Peng, T.L.; Reiders, S.T.
 Am. J. Hum. Genet. 49, 545-554, 1991
 A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
 A/Reference number: A39786; MUID:91353570; PMID:1882840
 A/Accession: A39786
 A/Molecule type: mRNA
 A/Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A/Cross-references: GB:S55790; NID:G234418; PID:BA819637.1; PID:G234419
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
 C/Genetics:
 A/Gene: GDB:COL4A3
 A/Cross-references: GDB:128351; OMIM:120070
 A/Map position: 2q36-2q37
 A/Introns: 1395/1, 1418/1, 1488/1, 1547/2, 1585/3, 1643/2 #status incomplete
 A/Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
 er associations in the interrupted helical domain (with disulfide and desmosine cross-li
 C/Function:
 A/Description: minor structural component of extracellular basement membrane in kidney
 A/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAR>
 F:29-42/Domain: amino-terminal nonhelical, NH1 <NH1>

F:43-1438/Region: interrupted helical
 F:791-793/Region: cell attachment (R-G-D) motif
 F:996-998/Region: cell attachment (R-G-D) motif
 F:1154-1156/Region: cell attachment (R-G-D) motif
 F:1306-1308/Region: cell attachment (R-G-D) motif
 F:1345-1347/Region: cell attachment (R-G-D) motif
 F:1432-1434/Region: cell attachment (R-G-D) motif
 F:1439-1670/Domain: carboxyl-terminal nonhelical, NCL <NCL>
 F:1451-1655/Domain: collagen IV carboxyl-terminal repeat <CTR>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>
 F:31.33.39.41.125.422.476.479.682.722.809.1387/Disulfide bonds: interchain #status pred:
 F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1460-1548.1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511.1616-1622/Disulfide bonds: #status predicted
 F:1570-1662.1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
 Query Match 84.5%; Score 93; DB 1; Length 1670;
 Best Local Similarity 94.4%; Pred. No. 1.4e-05;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNVNVCVCFASNDYS 19
 DB 1503 LFCNVNVCVCFASNDYS 1520
 RESULT 3
 S49488
 collagen alpha 3(IV) chain - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
 C/Accession: S49488
 R/Oberbaumer, I.
 A/Description: Cloning of the NCL domains fo the minor collagen IV chains of mouse via
 cells.
 A/Reference number: S49487
 A/Accession: S49488
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-161 <OBE>
 A/Cross-references: EMBL:X82205; NID:G559472; PID:CAA57689.1; PID:G559916
 C/Superfamily: collagen alpha 1(IV) chain
 Query Match 83.6%; Score 92; DB 2; Length 161;
 Best Local Similarity 88.9%; Pred. No. 2.9e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNVNVCVCFASNDYS 19
 DB 12 LFCNVNVCVCFASNDYS 29
 RESULT 4
 I48302
 collagen alpha 3(IV) chain - mouse (fragment)
 C/Species: Mus musculus (house mouse)
 C/Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
 C/Accession: I48302; S47278
 R/Miner, J.H.; Sames, J.R.
 J. Cell Biol. 127, 879-891, 1994
 A/Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
 A/Reference number: A54979; MUID:95050957; PMID:7962065
 A/Accession: I48302
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-246 <RES>
 A/Cross-references: EMBL:X35166; NID:G535197; PID:G535198
 C/Superfamily: collagen alpha 1(IV) chain
 Query Match 83.6%; Score 92; DB 2; Length 246;
 Best Local Similarity 88.9%; Pred. No. 4.1e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASNDYIS 19
 |||||:|||||
 Db 79 LFCNINNVNCFASNDYIS 96
 |||||:|||||
 RESULT 5
 B61228
 collagen alpha 1(IV) chain - rabbit (fragment)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
 C:Accession: B61228
 R:Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.
 Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991
 A:Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium
 A:Reference number: A61228; MUID:92010685; PMID:1717398
 A:Accession: B61228
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-258 <YAM>
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 83.6%; Score 92; DB 2; Length 258;
 Best Local Similarity 88.9%; Pred. No. 4.2e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASNDYIS 19
 |||||:|||||
 Db 92 LFCNINNVNCFASNDYIS 109
 |||||:|||||
 RESULT 6
 A39024
 collagen alpha 3(IV) chain - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
 C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
 R:Morrison, K.E.; Germino, G.G.; Reeders, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the alpha3(IV) chain of bovine basement membrane type IV collagen
 A:Reference number: A39024; MUID:91093146; PMID:1985905
 A:Accession: A39024
 A:Molecule type: mRNA
 A:Residues: 1-471 <MOR>
 A:Cross-references: EMBL:M63139; NID:g162886; PIDN:AAA62708.1; PID:g162887
 R:Butkowski, R.J.; Langeveld, J.P.M.; Wieselander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane type IV collagen
 A:Reference number: S18432; MUID:87222419; PMID:2438283
 A:Accession: S20672
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-244 <BUT>
 R:Saus, J.; Wieselander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
 A:Reference number: S17802; MUID:98330844; PMID:3417661
 A:Accession: S17802
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 265, 5466-5469, 1990
 A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type IV collagen
 A:Reference number: A35167; MUID:90202779; PMID:2318822
 A:Accession: A35167
 A:Molecule type: protein
 A:Residues: 236-258 <GUN>
 R:Gunwar, S.; Ballerster, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
 J. Biol. Chem. 266, 15318-15324, 1991
 A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncollagenous chain of type IV collagen
 A:Reference number: A39419; MUID:91332055; PMID:1869555
 A:Accession: C39419
 A:Molecule type: protein
 A:Residues: 236-255 <GU2>
 C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;
 F:1-238/Domain: collagenous (fragment) #status predicted <COL>
 F:239-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
 F:354-471/Domain: repeat NCI #status predicted <NCI1>
 F:354-471/Domain: repeat NCI #status predicted <NCI2>
 F:232,238/Modified site: hydroxyproline (Pro) #status experimental
 F:306-312,417-423/Disulfide bonds: #status predicted

Query Match 83.6%; Score 92; DB 2; Length 471;
 Best Local Similarity 88.9%; Pred. No. 6.9e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASNDYIS 19
 |||||:|||||
 Db 304 LFCNINNVNCFASNDYIS 321
 |||||:|||||
 RESULT 7
 CGH4B
 collagen alpha 1(IV) chain precursor - human
 N:Alternate names: procollagen alpha 1(IV) chain
 C:Species: Homo sapiens (man)
 C>Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 07-Dec-1999
 C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S02007; S39614; A02863; A
 R:Soiminen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
 J. Biol. Chem. 264, 13565-13571, 1989
 A:Title: Structural organization of the gene for the alpha-1 chain of human type IV co
 A:Reference number: S16876; MUID:89340433; PMID:2701944
 A:Accession: S16876
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1669 <SO11>
 A:Cross-references: EMBL:J04217; GB:J05039; NID:g180800; PIDN:AAA53098.1; PID:g180803
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
 R:Soiminen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
 J. Biol. Chem. 263, 17217-17220, 1988
 A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen a
 A:Reference number: A92690; MUID:89034231; PMID:3182844
 A:Accession: A32117
 A:Molecule type: DNA
 A:Residues: 1-28 <SO12>
 A:Cross-references: EMBL:J04217; NID:g180759; PIDN:AAA53097.1; PID:g553233
 R:Proesch, B.; Pollner, R.; Kuehn, K.
 EMBO J. 7, 2687-2695, 1988
 A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane
 A:Reference number: S02738; MUID:89030632; PMID:2846280
 A:Accession: S02738
 A>Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-6, 'L', 8-28 <POE>
 A:Cross-references: EMBL:X12784; NID:g30072
 R:Brazel, D.; Oberbaumer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.
 Eur. J. Biochem. 168, 529-536, 1987
 A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement m
 A:Reference number: S00048; MUID:88029471; PMID:3311751
 A:Accession: S00048
 A:Molecule type: mRNA
 A:Residues: 1-318, 'A', 320-944 <BRAL>
 A:Cross-references: EMBL:X05561; NID:g30066; PIDN:CAA29075.1; PID:g30067
 A:Accession: S25826
 A:Molecule type: protein
 A:Residues: 271-318, 'A', 320-554 <BRAL>
 R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
 Eur. J. Biochem. 152, 213-219, 1985
 A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7
 A:Reference number: A23115; MUID:86004708; PMID:4043082
 A:Accession: A23115
 A:Molecule type: protein
 A:Residues: 28-236, 'X', 239-240, 'K', 242-243 <GLA>
 A:Experimental source: placenta
 A>Note: the amino end of the mature form is blocked
 R:Soiminen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
 FEBS Lett. 225, 188-194, 1987

A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane (type IV)
 A:Reference number: S00207; MUID:88083584; PMID:3691802
 A:Accession: S00207
 A:Molecule type: mRNA
 A:Residues: 244-530 <S013>
 A:Cross-references: EMBL:Y00706; NID:929548; PIDN:CAA68698.1; PID:929549
 R:Ele, J. A.; Golbik, R.; Mann, K.; Kuehn, K.
 EMBO J. 12, 4795-4802, 1993
 A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
 A:Reference number: S39614; MUID:94038963; PMID:8223488
 A:Accession: S39614
 A:Molecule type: protein
 A:Residues: 371-554 <EBL>
 R:Babel, W.; Glanville, R. W.
 Eur. J. Biochem. 143, 545-556, 1984
 A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid sequence
 A:Reference number: A02863; MUID:85003629; PMID:6434307
 A:Accession: A02863
 A:Molecule type: protein
 A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-
 A:Experimental source: placenta
 R:Glanville, R. W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
 A:Title: Pepsin fragments of human placental basement-membrane collagens showing interrupted
 A:Reference number: S16908; MUID:82005835; PMID:6792033
 A:Accession: A58517
 A:Molecule type: protein
 A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553:1389-1405, 'XX', 1408-1409, 'X', 1411-14
 R:MacWright, R.; Benson, V. A.; Lovello, K. T.; van der Rest, M.; Fiezek, P. P.
 Biochemistry 22, 4940-4948, 1983
 A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (type
 A:Reference number: S16910; MUID:94053346; PMID:6416291
 A:Accession: S16910
 A:Molecule type: protein
 A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-549:939-940, 'M', 942-944, 'V', 946, 'X', 948-
 A:Experimental source: placenta
 R:Phinlan, T.; Tryggvason, K.; Myers, J. C.; Kurkinen, M.; Lebo, R.; Cheung, M. C.; F
 J. Biol. Chem. 260, 7681-7687, 1985
 A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
 A:Reference number: S01466; MUID:85207819; PMID:2581969
 A:Accession: S01466
 A:Molecule type: mRNA
 A:Residues: 1256-1669 <PTH>
 A:Cross-references: EMBL:M10940; NID:9180421; PIDN:AA52006.1; PID:g180424
 R:Brinker, J. M.; Gudas, L. J.; Loidl, H. R.; Wang, S. Y.; Rosenbloom, J.; Kefalides, N. A.;
 Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
 A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
 A:Reference number: S16879; MUID:85216555; PMID:2582422
 A:Accession: S16879
 A:Molecule type: mRNA
 A:Residues: 1259-1669 <BRI>
 A:Cross-references: EMBL:M11315; NID:g180817; PIDN:AA52042.1; PID:g180818
 R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A:Reference number: A02864; MUID:85127033; PMID:2578961
 A:Accession: S19091
 A:Molecule type: protein
 A:Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491:1501-1514, 'X', 1516-1519:1534-1553, 'X',
 R:Siebold, B.; Deutzmann, R.; Kuehn, K.
 Eur. J. Biochem. 176, 617-624, 1988
 A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
 A:Reference number: S02550; MUID:89005112; PMID:2844531
 A:Contents: annotation; disulfide bonds
 C:Genetics:
 A:Gene: GDB:COM4A1
 A:Cross-references: GDB:119791; OMIM:120130
 A:Map position: 13q34-13q34
 A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/
 1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
 C:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(
 C:Associations: among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
 r-trimer associations in the interrupted helical domain (with disulfide and desmosine cr

C:Function:
 A:Description: structural component of extracellular basement membrane
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicati
 F:1-36/Domain: signal sequence #status predicted <SIG>
 F:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
 F:29-162/Domain: amino-terminal nonhelical, 7S <7SD>
 F:163-1440/Domain: interrupted helical <COL>
 F:143-452/Region: integrin binding #status experimental
 F:597-599/Region: cell attachment (R-G-D) motif
 F:917-919/Region: cell attachment (R-G-D) motif
 F:968-970/Region: cell attachment (R-G-D) motif
 F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTI>
 F:127/Modified site: blocked amino end (Ara) (in mature form) #status experimental
 F:31-36,39,41,125,434,467,470/Disulfide bonds: interchain #status predicted
 F:45,48,78,90,129,156,217,228,231,277,295,298,322,343,361,460,463,497,527,540,543,
 1081,1084,1099,1117,1132,1150,1165,1182,1185,1188,1206,1235,1265,1283,1304,1319,1328,1
 F:45,48,78,90,129,156,217,228,231,277,295,298,322,343,361,460,463,497,527,543,573,582,
 99,1117,1132,1150,1165,1182,1185,1188,1206,1235,1265,1283,1304,1319,1328,1340,1356,137
 F:54,63,75,84,87,96,102,105,108,111,117,120,123,138,141,147,150,153,159,167,178,181,18
 F:49,422,425,439,445,448,451,479,485,491,494,503,512,518,524,530,546,549,552,555,561,5
 F:745,748,751,754,763/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:126/Binding site: carboxylate (Asn) (covalent) #status experimental
 F:129/Modified site: lysine (Lys) #status predicted
 F:172,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical
 F:272,645,839/Modified site: 4-hydroxyproline (Pro) #status atypical
 F:446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted
 F:766,775,784,787,790,796,804,810,816,822,834,860,863,869,872,875,887,890,893,899,
 23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,125
 F:1120,1268/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:1120,1268/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental
 F:1214,1424/Modified site: 3-hydroxyproline (Pro) #status absent
 F:1395,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
 Query Match 83.6%; Score 92; DB 1; Length 1669;
 Best Local Similarity 88.9%; Pred. No. 1.9e-05;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNNVCNCFASRNDYS 19
 |||||
 DB 1503 LFCNNVCNCFASRNDYS 1520
 RESULT 8
 CGMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000
 C:Accession: A33525; S01454; A28065; A02864; A25636; A29301; S19079; A32003; A31766; S
 R:Murthukumar, G.; Blumberg, B.; Kurkinen, M.
 J. Biol. Chem. 264, 6310-6317, 1989
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Di
 A:Reference number: A33525; MUID:89197932; PMID:2703490
 A:Accession: A33525
 A:Molecule type: mRNA
 A:Residues: 1-1669 <MT>
 A:Cross-references: EMBL:J04694; NID:9556296; PIDN:AA50292.1; PID:g556297
 R:Wood, L.; Theriault, N.; Vogeli, G.
 FEBS Lett. 227, 5-8, 1988
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the
 A:Reference number: S01454; MUID:88112221; PMID:3338568
 A:Accession: S01454
 A:Molecule type: mRNA
 A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-
 A:Cross-references: EMBL:X06777
 R:Killen, P. D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988

A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain
 A;Reference number: A28066; MUID:88243724; PMID:3379041
 A;Accession: A28066
 A;Molecule type: mRNA
 A;Residues: 1-129 <X1>
 A;Cross-references: EMBL:J03758; NID:G192669; PIDN:AAA37439.1; PID:G192670
 R;Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1985
 A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A;Reference number: A02864; MUID:85127033; PMID:2578961
 A;Accession: A02864
 A;Molecule type: mRNA
 A;Residues: 1276-1669 <OB>
 A;Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876
 R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G. Gene 43, 301-304, 1986
 A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
 A;Reference number: A25636; MUID:86301886; PMID:3755692
 A;Accession: A25636
 A;Molecule type: mRNA
 A;Residues: 1149-1396, S', 1398-1424 <NAT>
 A;Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287
 A;Note: the authors translated the codon CAG for residue 1374 as Arg
 R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj
 J. Biol. Chem. 262, 8496-8499, 1987
 A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
 A;Reference number: A94680; MUID:87250460; PMID:3597383
 A;Accession: A29301
 A;Molecule type: mRNA
 A;Residues: 1441-1669 <XUR>
 A;Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115
 R;Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y. J. Biol. Chem. 263, 12310-12314, 1988
 A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
 A;Reference number: S19079; MUID:88315019; PMID:2842328
 A;Accession: S19079
 A;Molecule type: DNA
 A;Residues: 1-28 <K2>
 A;Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503
 R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G. J. Biol. Chem. 263, 19274-19277, 1988
 A;Title: Head-to-head arrangement of murine type IV collagen genes.
 A;Reference number: A92702; MUID:89066738; PMID:3198626
 A;Accession: A32003
 A;Molecule type: DNA
 A;Residues: 1-28 <KAY>
 A;Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G504449
 R;Burbello, P.D.; Martin, G.R.; Yamada, Y. Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
 A;Reference number: A94220; MUID:89071759; PMID:3200851
 A;Accession: A31766
 A;Molecule type: DNA
 A;Residues: 1-28 <BUR>
 A;Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668
 R;Sakurai, Y.; Sullivan, M.; Yamada, Y. J. Biol. Chem. 261, 6654-6657, 1986
 A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
 A;Reference number: S19094; MUID:86196099; PMID:3009468
 A;Accession: S19094
 A;Molecule type: DNA
 A;Residues: 1110-1135; 1189-1336; 1342-1383; 1418-1487 <SAK>
 A;Cross-references: EMBL:M13027
 R;Schuppan, D.; Timpl, R.; Glanville, R.W. FEBS Lett. 115, 297-300, 1980
 A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
 A;Reference number: S16909; MUID:80246483; PMID:6772473
 A;Accession: S16909
 A;Molecule type: protein
 A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123
 R;Schuppan, D.; Glanville, R.W.; Timpl, R. Eur. J. Biochem. 123, 505-512, 1982
 A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amid

A;Reference number: A25991; MUID:82186723; PMID:6804236
 A;Accession: A25991
 A;Molecule type: protein
 A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 1
 61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-111
 A;Accession: B25991
 A;Molecule type: protein
 A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'O', 1207, 'XE', 1210-123
 3, 'SP', 1266, 'IV', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 13
 R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R. Eur. J. Biochem. 139, 401-410, 1984
 A;Title: Subunit structure and assembly of the globular domain of basement-membrane co
 A;Reference number: S17801; MUID:84132058; PMID:6698021
 A;Accession: S17801
 A;Molecule type: protein
 A;Residues: 1435-1443 <WEB>
 C;Genetics:
 A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
 A;Note: the list of introns may be incomplete
 C;Superfamily: collagen alpha 1(IV) chain
 C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular m
 F;1-27/Domain: signal sequence #status predicted <SIG>
 F;28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
 F;28-162/Domain: 7S <7SD>
 F;163-1440/Domain: collagenous, triple helix <COL>
 F;597-599/Region: cell attachment (R-G-D) motif
 F;781-783/Region: cell attachment (R-G-D) motif
 F;917-919/Region: cell attachment (R-G-D) motif
 F;968-970/Region: cell attachment (R-G-D) motif
 F;1441-1669/Domain: carboxyl-terminal nonhelical, NCI <NC1>
 F;1441-1552/Region: duplication
 F;1553-1669/Region: duplication
 F;31.36.39.41.434.467.470/Disulfide bonds: interchain #status predicted
 F;126/Banding site: carbohydrate (Asn) (covalent) #status predicted
 F;971,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,1090
 92,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxyp
 F;1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental
 F;1505-1511,1616-1622/Disulfide bonds: #status predicted
 Query Match 83.68; Score 92; DB 1; Length 1669;
 Best Local Similarity 88.98; Pred. No. 1.9e-05;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNVNVCNCFASNDYS 19
 DB 1503 LFCNINNVNCFASNDYS 1520
 RESULT 9
 I48304
 collagen alpha 5(IV) chain - mouse (fragment)
 C;Species: Mus musculus (house mouse)
 C;Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999
 C;Accession: I48304; S47280
 R;Miner, J.H.; Sames, J.R. J. Cell Biol. 127, 879-891, 1994
 A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq
 A;Reference number: A54979; MUID:95050957; PMID:7962065
 A;Accession: I48304
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-253 <RES>
 A;Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202
 C;Superfamily: collagen alpha 1(IV) chain
 Query Match 81.88; Score 90; DB 2; Length 253;
 Best Local Similarity 83.38; Pred. No. 8.1e-06;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNVNVCNCFASNDYS 19
 DB 87 MFCNINNVNCFASNDYS 104

RESULT 10

A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267
R:Zheng, K.; Thorner, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-1993, 1994
A>Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked type IV.
A:Reference number: A55267; MUID:94224868; PMID:8171024
A:Accession: A55267
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1754 <ZHE>
A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548
C:Superfamily: collagen alpha 1(IV) chain

| Query Match | 81.8% | Score 90; | DB 2; | Length 754; |
|-----------------------|-------|------------------|-------|-------------|
| Best Local Similarity | 83.3% | Pred. No. 2e-05; | | |
| Matches | 15; | Conservative | 2; | Mismatches |
| | | | 1; | Indels |
| | | | 0; | Gaps |

QY 2 LFCNVNVCVFASNDYS 19
:|||||
DB 595 MFCNINNVNVCVFASNDYS 612.

RESULT 11

S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N:Alternate names: procollagen alpha 5(IV) chain
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
R:Zhong, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A>Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identification of Alport syndrome patient.
A:Reference number: S22917; MUID:92316923; PMID:1352287
A:Accession: S22917
A:Molecule type: mRNA
A:Residues: 1-967 <ZHO>
A:Cross-references: GB:M90464; NID:G180826; PIDN:AAAS2046.1; PID:G553234
R:Zhong, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A>Title: Structure of the human type IV collagen COL4A5 gene.
A:Reference number: A54365; MUID:94165049; PMID:8120014
A:Accession: A54365
A:Molecule type: DNA
A:Residues: 1-922 <ZHD>
A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAC27816.1; PID
R:Zhong, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paeppe, A.; Tryggvason, K.
Science 261, 1167-1169, 1993
A>Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm
A:Reference number: A57079; MUID:93361972; PMID:8356449
A:Accession: A57079
A:Molecule type: DNA
A:Residues: 1-27 <ZHA>
A:Cross-references: GB:237153; NID:G587203; PIDN:CAA85512.1; PID:G587204
R:Piilajantemi, T.; Pohjola, E.R.; Myers, J.C.
J. Biol. Chem. 265, 13758-13766, 1990
A>Title: Complete primary structure of the triple-helical region and the carboxyl-termin
A:Reference number: A37122; MUID:90337990; PMID:2380186
A:Accession: A37122
A:Molecule type: mRNA
A:Residues: 84-439, 'GS', 442-624, 'LALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
A:Cross-references: GB:J05558; EMBL:M58526; NID:G1314209
A>Note: submitted to the EMBL Data Library, February 1991
A>Note: the authors translated the codon GCC for residue 115 as Val
R:Renieri, A.; Seri, M.; Myers, J.C.; Piilajantemi, T.; Massella, L.; Rizzoni, G.; De Ma

F;1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;129/Modified site: allysine (Iys) #status predicted

Query Match 72.7%; Score 80; DB 2; Length 1752;
Best Local Similarity 77.8%; Pred. No. 0.0011;
Matches 14; Conservative 2; Mismatches 0; Gaps 0;

QY 2 LFCNVNVCNFAASNDYS 19
|||||:|:|:|:|:|:|
Db 1586 LFCNNVCHVASRNDYS 1603

RESULT 13
AS4121
collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)
N;Alternate names: collagen alpha 2(IV) chain homolog
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
R;Accession: A54121; 944317
R;Exposito, J.Y.; Suzuki, H.; Geourjon, C.; Garrone, R.; Ramirez, F.
J. Biol. Chem. 269, 13167-13171, 1994
A;Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha 2
A;Reference number: A54121; MUID:94230414; PMID:8175744
A;Accession: A54121
A;Molecule type: mRNA
A;Residues: 1-1747 <EXP>
A;Cross-references: EMBL:X76730; NID:G483606; PIDN:CAA54146.1; PID:G483607
C;Genetics:
A;Gene: COLP4alpha
C;Superfamily: collagen alpha 1(IV) chain

Query Match 69.1%; Score 76; DB 2; Length 1747;
Best Local Similarity 77.8%; Pred. No. 0.0041;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFAASNDYS 19
|||||:|:|:|:|:|:|
Db 1584 LFCNFNNVCNYSRNDRS 1601

RESULT 14
SL6366
collagen alpha 2(IV) chain precursor - pig roundworm
C;Species: Ascaris suum (pig roundworm)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C;Accession: SL6366
R;Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the par
A;Reference number: SL6366; MUID:91340768; PMID:1714907
A;Accession: SL6366
A;Molecule type: mRNA
A;Residues: 1-1763 <JBI>
A;Cross-references: GB:M67507; NID:G159648; PIDN:AAA18014.1; PID:G159649
C;Genetics:
A;Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 556/3; 790/1; 891/1; 963/1
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulf
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F;27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F;43-1529/Domain: collagenous #status predicted <COL>
F;197-199/Region: cell attachment (R-G-D) motif
F;1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1530-1638/Domain: repeat NC1 #status predicted <NC11>
F;1639-1763/Domain: repeat NC1 #status predicted <NC12>
F;31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 69.1%; Score 76; DB 2; Length 1763;
Best Local Similarity 77.8%; Pred. No. 0.0041;

Mon Apr 5 07:53:16 2004

Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFSRNDYS 19
 |||:|||||:|||||
 Db 1591 LFCNVNVCNFSRNDYS 1608

RESULT 15

S40991
 Collagen alpha 1(IV) chain precursor - Caenorhabditis elegans
 N/Alternate names: protein K04H4.1
 C/Species: Caenorhabditis elegans
 C/Date: 03-May-1994 #sequence revision 02-Aug-1994 #text_change 13-Aug-1999
 C/Accession: S40991; S44442; S13651; B34476
 R/Ainscough, R.
 submitted to the EMBL Data Library, October 1993
 A/Reference number: S40991
 A/Accession: S40991
 A/Molecule type: DNA
 A/Residues: 1-1744 <AIN>
 A/Cross-references: EMBL:Z27078; NID:g414627; PID:g414628
 R/Kramer, J.M.
 submitted to the EMBL Data Library, December 1990
 A/Reference number: S44442
 A/Accession: S44442
 A/Molecule type: DNA
 A/Residues: 1-129, 'GFGMPGLAGPGQSGQNGNPGRLSGPPGGVNSQGRKGVKSGRSVGPLP', 209-281, 'PV
 15, 'D', 817-1260, 'P', 1262-1707, 'P', 1709-1744 <KRA>
 A/Cross-references: EMBL:X56979; NID:g6675; PIDN:CAA40299.1; PID:g6676
 R/Guo, X.; Johnson, J.J.; Kramer, J.M.
 Nature 349, 707-709, 1991
 A/Title: Embryonic lethality caused by mutations in basement membrane collagen of C. ele
 A/Reference number: S13651; MUID:91141582; PMID:1996137
 A/Accession: S13651
 A/Status: nucleic acid sequence not shown
 A/Molecule type: DNA
 A/Residues: 1-129, 'GFGMPGLAGPGQSGQNGNPGRLSGPPGGVNSQGRKGVKSGRSVGPLP', 209-281, 'PV
 15, 'D', 817-1260, 'P', 1262-1515 <GUI>
 A/Cross-references: EMBL:X56979
 R/Guo, X.; Kramer, J.M.
 J. Biol. Chem. 264, 17574-17582, 1989
 A/Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1
 A/Reference number: A34476; MUID:90008929; PMID:2793871
 A/Accession: B34476
 A/Molecule type: DNA
 A/Residues: 1432-1499, 'Q', 1501-1707, 'P', 1709-1744 <GU2>
 A/Cross-references: EMBL:J05067; NID:g156255; PIDN:AAB59179.1; PID:g156256
 C/Genetics:
 A/Gene: clb-2; emb-9
 A/Map position: 3
 A/Introps: 23/2; 79/1; 152/2; 288/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
 F:43-1515/Domain: collagenous, triple helix #status predicted <COL>
 F:93-95/Region: cell attachment (R-G-D) motif
 F:1053-1055/Region: cell attachment (R-G-D) motif
 F:1396-1398/Region: cell attachment (R-G-D) motif
 F:1516-1744/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
 F:1516-1627, 1628-1744/Region: duplication
 F:1580-1586, 1691-1697/Disulfide bonds: #status predicted

Query Match 68.2%; Score 75; DB 2; Length 1744;
 Best Local Similarity 66.7%; Pred. No. 0.0057;
 Matches 12; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFSRNDYS 19
 |||:|||||:|||||
 Db 1578 MFCNNSVCHVSRNDYS 1595

Search completed: April 5, 2004, 07:05:39
 Job time : 4.91041 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 2.39225 Seconds
(without alignments)
413.557 Million cell updates/sec

Title: US-10-032-221b-41

Perfect score: 110

Sequence: 1 KLFNCVNCVCFNSRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|--------------|---------------------|
| 1 | 93 | 84.5 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 2 | 92 | 83.6 | 471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 3 | 92 | 83.6 | 1669 | 1 CA14 HUMAN | P02462 homo sapien |
| 4 | 92 | 83.6 | 1669 | 1 CA14 MOUSE | P02462 mus musculus |
| 5 | 90 | 81.8 | 754 | 1 CA54 CANFA | Q28247 canis famil |
| 6 | 90 | 81.8 | 1685 | 1 CA54 HUMAN | P29400 homo sapien |
| 7 | 76 | 69.1 | 1763 | 1 CA24 ASCSU | P27393 ascaris suu |
| 8 | 75 | 68.2 | 1758 | 1 CA14 CAEEL | P17139 caenorhabdi |
| 9 | 67 | 60.9 | 1758 | 1 CA24 CAEEL | P17140 caenorhabdi |
| 10 | 64 | 58.2 | 1691 | 1 CA64 HUMAN | Q14031 homo sapien |
| 11 | 59 | 53.6 | 623 | 1 CA44 RABIT | P55787 oryctolagus |
| 12 | 59 | 53.6 | 1690 | 1 CA44 HUMAN | P53420 homo sapien |
| 13 | 59 | 53.6 | 1775 | 1 CA14 DROME | P08120 drosophila |
| 14 | 58 | 52.7 | 453 | 1 CA44 BOVIN | Q29442 bos taurus |
| 15 | 58 | 52.7 | 1707 | 1 CA24 MOUSE | P08122 mus musculus |
| 16 | 58 | 52.7 | 1712 | 1 CA24 HUMAN | P08572 homo sapien |
| 17 | 47 | 42.7 | 1477 | 1 HTK7 HYDAT | Q25197 hydra atten |
| 18 | 44 | 40.0 | 333 | 1 AMR1 HUMAN | Q95490 homo sapien |
| 19 | 44 | 40.0 | 344 | 1 AMR1 MOUSE | Q95490 mus musculus |
| 20 | 42.5 | 38.6 | 553 | 1 VH65 NPVAC | Q08539 autographa |
| 21 | 42.5 | 38.6 | 578 | 1 VAC8 YEAST | P39968 saccharomyc |
| 22 | 42 | 38.2 | 288 | 1 SPY3 HUMAN | Q43610 homo sapien |
| 23 | 42 | 38.2 | 312 | 1 XYN1 CALSA | P23557 caldocellum |
| 24 | 42 | 38.2 | 339 | 1 OTC EUCBP | Q89891 buchnera ap |
| 25 | 42 | 38.2 | 461 | 1 TR1B HUMAN | P20333 homo sapien |
| 26 | 42 | 38.2 | 474 | 1 TR1B MOUSE | P25119 mus musculus |
| 27 | 42 | 38.2 | 589 | 1 SPY DROME | Q44783 drosophila |
| 28 | 42 | 38.2 | 703 | 1 NH55 CAEEL | Q9na51 caenorhabdi |
| 29 | 42 | 38.2 | 1216 | 1 ATU1 YEAST | P38360 saccharomyc |
| 30 | 42 | 38.2 | 6629 | 1 RIAB IBVBC | P27920 a replicase |
| 31 | 42 | 38.2 | 6629 | 1 RIAB IBVBC | Q91qt2 a replicase |
| 32 | 41 | 37.3 | 60 | 1 IT11 BRANA | P80301 brassica na |
| 33 | 41 | 37.3 | 184 | 1 MTR2 YEAST | P34232 saccharomyc |

RESULT 1

CA34 HUMAN
ID CA34 HUMAN STANDARD; PRT; 1670 AA.
AC Q01955; O9BQT2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain;
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. Biol. Chem. 269:23013-23017(1994).
RN [2]
RP REVISIONS.
RA Leinonen A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
RP CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-405; ARG-451;
RP PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064696; PubMed=11134255;
RA Heidet L., Arondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RJ J. Am. Soc. Nephrol. 12:97-106(2001).
RN [4]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RJ J. Biol. Chem. 267:19780-19784(1992).
RN [5]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=93353570; PubMed=1882840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RJ Am. J. Hum. Genet. 49:545-554(1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;

34 41 37.3 292 1 EFTS RALSO Q8xzj0 ralstonia s
35 41 37.3 313 1 EFTS ANASP Q8vny3 anabaena sp
36 41 37.3 327 1 CPZ7_PIG P79492 sus scrofa
37 41 37.3 365 1 H182 BRAJA Q89ul9 bradyrhizob
38 41 37.3 480 1 CP21 RABIT P00180 oryctolagus
39 41 37.3 489 1 CP26 CANFA O62671 canis famil
40 41 37.3 490 1 CP22 RABIT P00181 oryctolagus
41 41 37.3 490 1 CPC2 MSAU P33264 mesocricetu
42 41 37.3 506 1 Y619 METJA Q58036 methanococc
43 41 37.3 567 1 ARS STRPU P50473 strongyloce
44 41 37.3 623 1 PTR2 CANAL P46030 candida alb
45 41 37.3 2476 1 ZAN_PIG Q28983 sus scrofa

ALIGNMENTS

CC [3] unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26576; AAA53098.1; JOINED.
CC EMBL; J04217; AAA53098.1; JOINED.
CC EMBL; M26550; AAA53098.1; JOINED.
CC EMBL; M26540; AAA53098.1; JOINED.
CC EMBL; M26542; AAA53098.1; JOINED.
CC EMBL; M26543; AAA53098.1; JOINED.
CC EMBL; M26544; AAA53098.1; JOINED.
CC EMBL; M26545; AAA53098.1; JOINED.
CC EMBL; M26546; AAA53098.1; JOINED.
CC EMBL; M26547; AAA53098.1; JOINED.
CC EMBL; M26537; AAA53098.1; JOINED.
CC EMBL; M26538; AAA53098.1; JOINED.
CC EMBL; M26548; AAA53098.1; JOINED.
CC EMBL; M26549; AAA53098.1; JOINED.
CC EMBL; M26551; AAA53098.1; JOINED.
CC EMBL; M26552; AAA53098.1; JOINED.
CC EMBL; M26553; AAA53098.1; JOINED.
CC EMBL; M26554; AAA53098.1; JOINED.
CC EMBL; M26555; AAA53098.1; JOINED.
CC EMBL; M26556; AAA53098.1; JOINED.
CC EMBL; M26557; AAA53098.1; JOINED.
CC EMBL; M26539; AAA53098.1; JOINED.
CC EMBL; M26558; AAA53098.1; JOINED.
CC EMBL; M26559; AAA53098.1; JOINED.
CC EMBL; M26560; AAA53098.1; JOINED.
CC EMBL; M26561; AAA53098.1; JOINED.
CC EMBL; M26562; AAA53098.1; JOINED.
CC EMBL; M26563; AAA53098.1; JOINED.
CC EMBL; M26564; AAA53098.1; JOINED.
CC EMBL; M26565; AAA53098.1; JOINED.
CC EMBL; M26566; AAA53098.1; JOINED.
CC EMBL; M26567; AAA53098.1; JOINED.
CC EMBL; M26568; AAA53098.1; JOINED.
CC EMBL; M26569; AAA53098.1; JOINED.
CC EMBL; M26570; AAA53098.1; JOINED.
CC EMBL; M26571; AAA53098.1; JOINED.
CC EMBL; M26572; AAA53098.1; JOINED.
CC EMBL; M26573; AAA53098.1; JOINED.
CC EMBL; M26574; AAA53098.1; JOINED.
CC EMBL; M26575; AAA53098.1; JOINED.
CC EMBL; Y00706; CAA69698.1; --
CC EMBL; X05561; CAA29075.1; --
CC EMBL; M10940; AAA52006.1; --
CC EMBL; M11315; AAA52042.1; --
CC PTR; S16876; CGHU4B.
CC Genew; HGNC:2202; COL4A1.
CC MW; 120130; --
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagn4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 24.
CC ProDom; PD000007; C1g_helix; 6.
CC ProDom; PD003923; ProcollagnC4; 1.
CC SMART; SM00111; C4; 2.
RN RP SEQUENCE OF 1-943 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=86029471; PubMed=3311751;
RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,
RA Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
RT basement membrane collagen (type IV) reveals 21 non-triplet
RT interruptions located within the collagenous domain.";
RL Eur. J. Biochem. 168:529-536(1987).
RN [4]
RN RP SEQUENCE OF 28-243.
RX MEDLINE=86004708; PubMed=4043082;
RA Glanville R.W., Ghan R.O., Siebold B., Risteli J., Kuehn K.;
RT "Amino acid sequence of the N-terminal aggregation and cross-linking
RT region (7S domain) of the alpha 1 (IV) chain of human basement
RT membrane collagen.";
RL Eur. J. Biochem. 152:213-219(1985).
RN [5]
RN RP SEQUENCE OF 534-1447.
RX MEDLINE=8503629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RT "Structure of human-basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1(IV) chain.";
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RN RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85207819; PubMed=2581969;
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA Cheung M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RN RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RN RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soiminen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RN RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyl-terminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -1- FUNCTION: type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Lysines at the third position of the tripeptide repeating

KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NON-HELICAL REGION (NC1).
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
 FT DSULFID 1460 1551 OR 1548.
 FT DSULFID 1493 1548 OR 1551.
 FT DSULFID 1505 1511
 FT DSULFID 1570 1665 OR 1662.
 FT DSULFID 1604 1662 OR 1665.
 FT DSULFID 1616 1622
 FT CONFLICT 237 238 SG -> KE (IN REF. 4).
 FT CONFLICT 241 241 G -> K (IN REF. 4).
 FT CONFLICT 319 319 Q -> A (IN REF. 3).
 FT CONFLICT 719 719 N -> D (IN REF. 5).
 FT CONFLICT 837 837 D -> Y (IN REF. 5).
 FT CONFLICT 842 842 K -> P (IN REF. 5).
 FT CONFLICT 896 896 V -> W (IN REF. 2).
 FT CONFLICT 904 904 E -> Q (IN REF. 5).
 FT CONFLICT 914 914 S -> K (IN REF. 5).
 FT CONFLICT 998 998 S -> K (IN REF. 5).
 FT CONFLICT 1010 1010 K -> P (IN REF. 5).
 FT CONFLICT 1012 1012 S -> K (IN REF. 5).
 FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
 SQ SEQUENCE 1669 AA; 160611 MW; 3BEBAGDFFB9B8A4 CRC64;
 Query Match 83.6%; Score 92; DB 1; Length 1669;
 Best Local Similarity 88.9%; Pred. No. 3.4e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 LFCNVNVCNFCASRNDYS 19
 Db 1503 LFCNVNVCNFCASRNDYS 1520
 RESULT 4
 ID CA14_MOUSE STANDARD; PRT; 1669 AA.
 AC P02463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=89197932; PubMed=2703490;
 RA Muthukumar G., Blumberg B., Kurkinen M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 collagen IV. Differential evolution of collagen IV domains.";
 RL J. Biol. Chem. 264:6310-6317(1989).
 RN [2]
 RN SEQUENCE OF 1-1154 FROM N.A.
 RX MEDLINE=89112221; PubMed=3338568;
 RA Wood L., Theriault N., Vogeli G.;
 RT "cDNA clones completing the nucleotide and derived amino acid
 sequence of the alpha 1 chain of basement membrane (type IV) collagen
 from mouse.";
 RL FEBS Lett. 227:5-8(1988).
 RN [3]
 RN SEQUENCE OF 1149-1424 FROM N.A.
 RX MEDLINE=86301886; PubMed=3755692;
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
 synthetic oligodeoxynucleotide.";
 RL Gene 43:301-304(1986).
 RN [4]
 RN SEQUENCE OF 1276-1669 FROM N.A.
 RX MEDLINE=85127033; PubMed=2578961;
 RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
 RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
 the alpha 1(IV) chain of basement membrane collagen as derived from
 complementary DNA.";
 RL Eur. J. Biochem. 147:217-224(1985).
 RN [5]
 RN SEQUENCE OF 1441-1669 FROM N.A.
 RX MEDLINE=87250460; PubMed=3597383;
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
 RA Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse
 alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8496-8499(1987).
 RN [6]
 RN PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=86196099; PubMed=3009468;
 RA Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
 collagen genes.";
 RL J. Biol. Chem. 261:6654-6657(1986).
 RN [7]
 RN SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89066738; PubMed=3198626;
 RA Kayes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
 RT "Head-to-head arrangement of murine type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277(1988).
 RN [8]
 RN SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89071759; PubMed=3200851;
 RA Burdello P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
 bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
 RN [9]
 RN SEQUENCE OF 1-129 FROM N.A.
 RX MEDLINE=88243724; PubMed=3379041;
 RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
 collagen chain and the corresponding region of the gene.";
 RL J. Biol. Chem. 263:8706-8709(1988).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 CC alpha 6(IV), each of which can form a triple helix structure with
 CC 2 other chains to generate type IV collagen network.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-
 CC X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03758; AAA37439.1; -
 CC EMBL: M23333; AAA51625.1; -

```
DR EMBL: J04694; AAA50292.1; -
DR EMBL: X06777; CAA29945.1; -
DR EMBL: X02201; CAA26132.1; -
DR EMBL: M15832; AAA37342.1; -
DR EMBL: M14042; AAA37342.1; -
DR EMBL: M12879; AAA37343.1; -
DR EMBL: M13024; -; NOT_ANNOTATED_CDS.
DR EMBL: M13025; -; NOT_ANNOTATED_CDS.
DR EMBL: M13026; AAA37344.1; -
DR EMBL: M13027; AAA37345.1; -
DR EMBL: M13043; AAA37346.1; -
DR EMBL: J04448; AAA37437.1; -
DR PIR: A33525; CGMS4B.
DR MGD; MG1:88454; Col1a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg helix; 6.
DR ProDom; PD001923; Procollagn4_C.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1511 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 BY SIMILARITY.
FT DISULFID 1616 1622 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 126 126 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 83.6%; Score 92; DB 1; Length 1669;
Best Local Similarity 89.9%; Pred. No. 3.4e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASRDYS 19
DQ 1503 LFCNINNVNCFASRDYS 1520
|||||:|||||
|||||:|||||

RESULT 5
CA54_CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; TISSUE=Kidney;
RA MEDLINE=34224888; PubMed=8171024;
RX Zheng K., Thorner P.S., Marrano P., Bauml R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
CC canine X-linked hereditary nephritis (HN), a disease similar to
CC that in humans (also referred to as Alport syndrome) characterized
CC by progressive renal failure and neurosensory deafness.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: U07888; AAB60258.1; -.
DR PIR; A55267; A55267. Clg helix.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; Clg helix; 1.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 1
FT DOMAIN <1 530 TRIPLE-HELICAL REGION.
FT DOMAIN 531 >754 NONHELICAL REGION (NC1).
FT DISULFID 552 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 ? OR 754 (BY SIMILARITY).
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON_TER 754 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 81.8%; Score 90; DB 1; Length 754;
Best Local Similarity 83.3%; Pred. No. 3.1e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASRDYS 19
DQ 595 MFCNINNVNCFASRDYS 612
|||||:|||||
|||||:|||||
```

RESULT 6

CA54 HUMAN
ID CA54 HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Syers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;

RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Descheres G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
RN [14]
RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796; ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
RN [15]
RP VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M., Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
RN [16]
RP VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Brutini M., Galli L., Zanelli P., Neri T.M., Rossetti S., Tuccillo A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Baufi G., Scolarli F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F., Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51 exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).

```

[17]
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=970941179; PubMed=8940267;
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
RA Glatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
[18]
RP VARIANTS AS ASP-1498.
RP MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobrykina V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of  $\alpha 5$ (IV) collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
[19]
RP VARIANTS AS GLN-1677.
RP MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677O.";
RL Hum. Genet. 99:681-684(1997).
[20]
RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciari A., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
[21]
RP VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP 802-GLY--GLU-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
RX MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,
RA Sprangate J., Shows T.B., Petersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
[22]
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
RP SER-1170 AND TRP-1678; AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
RT sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
[23]
RP VARIANTS AS ARG-822.
Query Match 81.8%; Score 90; DB 1; Length 1685;
Best Local Similarity 83.3%; Pred. No. 6.8e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LFCNVNVCNPFASNDYS 19
Db 1519 MFCNINVCNPFASNDYS 1536
:||||:|||||
RESULT 7
CA24_ACSU
ID CA24_ACSU PRT; 1763 AA.
AC P27393;
DT 01-AUG-1992 (Rel. 23, Created)
```

```
FT isoform II).
FT /FTID=VSP 001159,
SQ SEQUENCE 1763 AA; 168526 MW; 304F5288C06A80D CRC64;
Query Match 69.1%; Score 76; DB 1; Length 1763;
Best Local Similarity 77.8%; Pred. No. 0.00094;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 LFCNVNVCNFASENNDYS 19
DB 1591 LFCNVNVCNFASENNDYS 1608
RESULT 8
CA14 CAEL STANDARD; PRT; 1758 AA.
AC P17139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CLB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
RL collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Karshaw J., Kitsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RL genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
```


CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eye, and Kidney;
RX MEDLINE=9417179; PubMed=8125972;
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;
RT Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RL J. Biol. Chem. 269:7520-7526 (1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Reders S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199 (1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Reders S.T., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grafham D., Lawlor S., Wilson S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=8356449;
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,
RA de Paeppe A., Tryggvason K., Reders S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169 (1993).
CC !- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC network together with laminins, proteoglycans and entactin/
CC nidogen.
CC !- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC !- SUBCELLULAR LOCATION: Cell surface (Potential).
CC !- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequences=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequences=VSP 001174;
CC !- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC !- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC !- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC !- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D21337; BAA04809.1; --
DR EMBL; U04845; AAL19569.2; --
DR EMBL; U47004; AAB19038.1; --
DR EMBL; U46959; AAB19038.1; JOINED.
DR EMBL; U46961; AAB19038.1; JOINED.
DR EMBL; U46962; AAB19038.1; JOINED.
DR EMBL; U46963; AAB19038.1; JOINED.
DR EMBL; U46964; AAB19038.1; JOINED.
DR EMBL; U46965; AAB19038.1; JOINED.
DR EMBL; U46966; AAB19038.1; JOINED.
DR EMBL; U46967; AAB19038.1; JOINED.
DR EMBL; U46968; AAB19038.1; JOINED.
DR EMBL; U46969; AAB19038.1; JOINED.
DR EMBL; U46970; AAB19038.1; JOINED.
DR EMBL; U46971; AAB19038.1; JOINED.
DR EMBL; U46972; AAB19038.1; JOINED.
DR EMBL; U46973; AAB19038.1; JOINED.
DR EMBL; U46974; AAB19038.1; JOINED.
DR EMBL; U46975; AAB19038.1; JOINED.
DR EMBL; U46976; AAB19038.1; JOINED.
DR EMBL; U46977; AAB19038.1; JOINED.
DR EMBL; U46978; AAB19038.1; JOINED.
DR EMBL; U46979; AAB19038.1; JOINED.
DR EMBL; U46980; AAB19038.1; JOINED.
DR EMBL; U46981; AAB19038.1; JOINED.
DR EMBL; U46982; AAB19038.1; JOINED.
DR EMBL; U46983; AAB19038.1; JOINED.
DR EMBL; U46984; AAB19038.1; JOINED.
DR EMBL; U46985; AAB19038.1; JOINED.
DR EMBL; U46986; AAB19038.1; JOINED.
DR EMBL; U46987; AAB19038.1; JOINED.
DR EMBL; U46988; AAB19038.1; JOINED.
DR EMBL; U46989; AAB19038.1; JOINED.
DR EMBL; U46990; AAB19038.1; JOINED.
DR EMBL; U46991; AAB19038.1; JOINED.
DR EMBL; U46992; AAB19038.1; JOINED.
DR EMBL; U46993; AAB19038.1; JOINED.
DR EMBL; U46994; AAB19038.1; JOINED.
DR EMBL; U46995; AAB19038.1; JOINED.
DR EMBL; U46996; AAB19038.1; JOINED.
DR EMBL; U46997; AAB19038.1; JOINED.
DR EMBL; U46998; AAB19038.1; JOINED.
DR EMBL; U46999; AAB19038.1; JOINED.
DR EMBL; U47000; AAB19038.1; JOINED.
DR EMBL; U47001; AAB19038.1; JOINED.
DR EMBL; U47002; AAB19038.1; JOINED.
DR EMBL; U47003; AAB19038.1; JOINED.
DR EMBL; U47004; AAB19039.1; --
DR EMBL; U46960; AAB19039.1; JOINED.
DR EMBL; U46961; AAB19039.1; JOINED.
DR EMBL; U46962; AAB19039.1; JOINED.
DR EMBL; U46963; AAB19039.1; JOINED.
DR EMBL; U46964; AAB19039.1; JOINED.
DR EMBL; U46965; AAB19039.1; JOINED.
DR EMBL; U46966; AAB19039.1; JOINED.
DR EMBL; U46967; AAB19039.1; JOINED.
DR EMBL; U46968; AAB19039.1; JOINED.
DR EMBL; U46969; AAB19039.1; JOINED.
DR EMBL; U46970; AAB19039.1; JOINED.
DR EMBL; U46971; AAB19039.1; JOINED.
DR EMBL; U46972; AAB19039.1; JOINED.
DR EMBL; U46973; AAB19039.1; JOINED.
DR EMBL; U46974; AAB19039.1; JOINED.
DR EMBL; U46975; AAB19039.1; JOINED.
DR EMBL; U46976; AAB19039.1; JOINED.
DR EMBL; U46977; AAB19039.1; JOINED.
DR EMBL; U46978; AAB19039.1; JOINED.
DR EMBL; U46979; AAB19039.1; JOINED.
DR EMBL; U46980; AAB19039.1; JOINED.
DR EMBL; U46981; AAB19039.1; JOINED.
DR EMBL; U46982; AAB19039.1; JOINED.
DR EMBL; U46983; AAB19039.1; JOINED.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=95014445; PubMed=7523402;
RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human type IV collagen alpha 4(IV)
RT chain. Comparison with structure and expression of the other alpha
RT (IV) chains.";
RL J. Biol. Chem. 269:26172-26177(1994).
RN [2]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Monota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [3]
RP SEQUENCE OF 1219-1690 FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=93374047; PubMed=8365481;
RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
RT "cDNA isolation and partial gene structure of the human alpha 4(IV)
RT collagen chain.";
RL FEBS Lett. 330:122-128(1993).
RN [4]
RP SEQUENCE OF 1407-1507 FROM N.A.
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
RN [5]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [6]
RP VARIANT AS SER-1201.
RX MEDLINE=95078927; PubMed=7987396;
RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
RA Smeets H.J.M., Reiders S.T.;
RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
RT collagen genes in autosomal recessive Alport syndrome.";
RL Nat. Genet. 8:77-82(1994).
RN [7]
RP VARIANT FBH GLU-897.
RX MEDLINE=96379660; PubMed=8787673;
RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
RT "Benign familial hematuria due to mutation of the type IV collagen
RT alpha4 gene.";
RL J. Clin. Invest. 98:1114-1118(1996).
RN [8]
RP VARIANTS AS AND VARIANTS.
RX MEDLINE=99011253; PubMed=9792860;
RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
RA Cochot P., Gruenfeld J.-P., Falcoix J.-B., Gubler M.-C., Antignac C.;
RT "Determination of the genomic structure of the COL4A4 gene and of
RT novel mutations causing autosomal recessive Alport syndrome.";
RL Am. J. Hum. Genet. 63:1329-1340(1998).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -1- DISEASE: Defects in COL4A4 are a cause of autosomal recessive
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder
CC characterized by progressive glomerulonephritis, renal failure,
CC hematuria, ocular abnormalities and deafness. The recessive form
CC occurs equally between males and females.
CC -1- DISEASE: Defects in COL4A4 are a cause of familial benign
CC hematuria (FBH) [MIM:141200] or thin basement membrane disease.
CC FBH is characterized by persistent hematuria, an electron
CC microscopically detectable thin glomerular basement membrane (GBM)
CC and an autosomal dominant mode of inheritance. Renal function
CC remains normal. In children, differentiation between FBH and AS
CC can be difficult, because both disorders are manifested by
CC persistent hematuria and thin GBM at that age.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X81053; CAA56943.1; --
CC EMBL; AB008496; BAA25065.1; --
CC EMBL; D17391; BAA04214.1; --
CC PIR; A55360; CGHUIB.
CC Genew; HGNC:2206; COL4A4.
CC MIM; 120131; --
CC MIM; 141200; --
CC MIM; 203780; --
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollag4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 21.
CC ProDom; PD000007; C1g_helix; 3.
CC ProDom; PD003923; Procollag4; 1.
CC SMART; SM00111; C4; 2.
CC Extracellular matrix; Connective tissue; Basement membrane; Repeat;
CC Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
CC Polymorphism; Alport syndrome.
FT SIGNAL 1 38
FT CHAIN 39 1690
FT COLLAGEN ALPHA 4(IV) CHAIN.
FT DOMAIN 39 64
FT 7S DOMAIN.
FT DOMAIN 65 1459
FT TRIPLE-HELICAL REGION.
FT DOMAIN 1460 1690
FT NONHELICAL REGION (NC1).
FT SITE 94 96
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 145 147
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 189 191
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 310 312
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 724 726
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 785 787
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 989 991
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1206 1207
FT CLEAVAGE (BY COLLAGENASE)
FT (BY SIMILARITY).

```

FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
FT DISULFID 1525 1531 BY SIMILARITY.
FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
FT DISULFID 1634 1641 BY SIMILARITY.
FT CARBOHYD 1432 1442 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLNAC. . .) (POTENTIAL).
FT VARIANT 441 446 Missing (in AS).
FT VARIANT 545 545 G -> A (in dBSNP:1800516).
FT VARIANT 570 570 E -> Q.
FT VARIANT 897 897 /FTID=VAR_008150.
FT VARIANT 931 931 /FTID=VAR_001912.
FT VARIANT 1004 1004 A -> T.
FT VARIANT 1030 1030 L -> P (in dBSNP:1800517).
FT VARIANT 1201 1201 /FTID=VAR_008152.
FT VARIANT 1402 1402 G -> V (in AS).
FT VARIANT 1572 1572 /FTID=VAR_008153.
FT VARIANT 1659 1660 G -> S (in AS).
FT CONFLICT 1690 AA; 164095 MW; E1E72F293A72BAAE CRC64;
SQ SEQUENCE 1690 AA; 164095 MW; E1E72F293A72BAAE CRC64;

Query Match 53.6%; Score 59; DB 1; Length 1690;
Best Local Similarity 52.9%; Pred. No. 0.34;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 FCNVNVCNPFASNDYS 19
Db 1524 YCNIHQVCHYAORNDRS 1540

RESULT 13
CA14_DROME STANDARD; PRT; 1775 AA.
AC P08120;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN C25C OR DCG1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89054012; PubMed=3142875;
RA Blumberg B., Mackrell A.J., Fessler J.H.;
RA "Drosophila basement membrane procollagen alpha 1(IV). II. Complete
RT cDNA sequence, genomic structure, and general implications for
RT supramolecular assemblies."
RL J. Biol. Chem. 263:18328-18337(1988).
RN [2]
RN SEQUENCE FROM N.A.
RP PROPEP 24 ? AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
RA Blumberg B.;
RA Thesis (1987), University of California / Los Angeles, U.S.A.
RN [3]
RN SEQUENCE FROM N.A.
RP PROPEP 24 ? AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
RA Blumberg B.;
RA Thesis (1987), University of California / Los Angeles, U.S.A.
RN [4]
RN SEQUENCE OF 1065-1775 FROM N.A.
RP DISULFID 1708 1767 OR 1770 (BY SIMILARITY).

```

```

PX MEDLINE=87194801; PubMed=3106346;
RA Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,
RA Natzle J.E., Fessler J.H.;
RA "Basement membrane procollagen IV and its specialized carboxyl domain
RT are conserved in Drosophila, mouse, and human."
RL J. Biol. Chem. 262:5947-5950(1987).
RN [5]
RN SEQUENCE OF 1355-1775 FROM N.A.
RX MEDLINE=87246644; PubMed=3109906;
RA Cecchini J.P., Knibiehler B., Mirre C., le Parco Y.;
RA "Evidence for a type-IV-related collagen in Drosophila melanogaster.
RT Evolutionary constancy of the carboxyl-terminal noncollagenous
RT domain."
RL Eur. J. Biochem. 165:587-593(1987).
RN [6]
RN SEQUENCE OF 762-1230 FROM N.A.
RX MEDLINE=82197577; PubMed=6210912;
RA Monson J.M., Natzle J., Friedman J., McCarthy B.J.;
RA "Expression and novel structure of a collagen gene in Drosophila."
RL Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23704; AAA28404.1; -
DR EMBL; M96575; AAB59184.1; -
DR EMBL; J02727; AAA28423.1; -
DR EMBL; M28334; AAA28422.1; -
DR EMBL; V00200; CAA23486.2; -
DR FIR; A31893; A31893.
DR FlyBase; FBgn0000299; Cg25C.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR InterPro; IPR008161; C1g Helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR01442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD000007; C1g_helix; 9.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 ? AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN ? 1775 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN ? 1544 TRIPLE-HELICAL REGION.
FT DOMAIN 1545 1775 NONHELICAL REGION (NC1).
FT DISULFID 1549 1655 OR 1652 (BY SIMILARITY).
FT DISULFID 1599 1652 OR 1655 (BY SIMILARITY).
FT DISULFID 1611 1617 BY SIMILARITY.
FT DISULFID 1674 1770 OR 1767 (BY SIMILARITY).
FT DISULFID 1708 1767 OR 1770 (BY SIMILARITY).

```

```
FT DISULFID 1720 1727 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC... ) (PROBABLE).
FT CONFLICT 948 948 L -> S (IN REF. 6).
FT CONFLICT 997 997 S -> T (IN REF. 6).
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).
FT CONFLICT 1373 1373 T -> I (IN REF. 5).
FT CONFLICT 1496 1496 L -> R (IN REF. 5).
FT CONFLICT 1507 1511 EYGNV -> RAGR (IN REF. 5).
FT CONFLICT 1529 1529 E -> K (IN REF. 5).
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AB23149525CD CRC64;

Query Match 53.6%; Score 59; DB 1; Length 1775;
Best Local Similarity 68.8%; Pred. No. 0.36;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRND 17
| | | | |
Db 1609 LSCGQNNVCNYSRND 1624

RESULT 14
CA44_BOVIN
ID CA44_BOVIN STANDARD; PRT; 453 AA.
AC Q29442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 4 (IV) chain (fragment).
GN COL4A4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Lens;
RX MEDLINE=92112769; PubMed=1370461;
RA Mariyama M., Kalluri R., Hudson B.G., Redders S.T.;
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4 (IV) and comparison with other type IV
RT collagens."
RL J. Biol. Chem. 267:1253-1258(1992).
RN [2]
RP SEQUENCE OF 217-246.
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen."
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 217-233.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen."
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
```

RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
R Deutmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain.";
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
RT "cDNA and protein sequence of the NCI domain of the alpha 2-chain of collagen IV and its comparison with alpha 1(IV).";
RL FEBS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
R Saus J., Fihlajantiemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";
RL FEBS Lett. 206:29-32(1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";
RL Nature 317:177-179(1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -|- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.
CC -|- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.
CC -|- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NCI) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.
CC -|- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -|- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NCI domain, are conserved in all known type IV collagens.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M23334; AAA51626.1; -
DR EMBL; M23333; AAA51626.1; JOINED.
DR EMBL; J04695; AAA50293.1; -
DR EMBL; J04448; AAA37438.1; -
DR EMBL; X04647; CAA28308.1; -
DR EMBL; M15833; AAA37341.1; -
DR EMBL; X04410; CAA27998.1; -
DR EMBL; X02896; CAA26655.1; -
DR EMBL; X02897; CAA51614.1; -
DR EMBL; X02898; CAA26657.1; -
DR EMBL; X02899; CAA26658.1; -
DR PIR; A33526; A33526.
DR MGD; MGI-88455; Col4a2.
DR GO; GO:0005504; C:Basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg_helix; 7.
DR ProDom; PD003923; Procollagen4; 1.
DR SMART; SM0111; C4; 2.
DR KX Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.
FT CHAIN 184 1707 TRIPLE-HELICAL REGION.
FT DOMAIN 184 1479 NONHELICAL REGION (NCI).
FT DOMAIN 1480 1707 OR 1585 (BY SIMILARITY).
FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).
FT DISULFID 1532 1585 BY SIMILARITY.
FT DISULFID 1544 1550 OR 1700 (BY SIMILARITY).
FT DISULFID 1607 1703 OR 1703 (BY SIMILARITY).
FT DISULFID 1641 1700 BY SIMILARITY.
FT DISULFID 1653 1660 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1270 1270 P -> R (IN REF. 6).
FT CONFLICT 1051 1051 S -> G (IN REF. 7).
FT CONFLICT 1097 1097 G -> S (IN REF. 6).
FT CONFLICT 1171 1171 P -> R (IN REF. 6).
FT CONFLICT 1179 1179 Q -> E (IN REF. 6).
FT CONFLICT 1241 1241 P -> A (IN REF. 6).
FT CONFLICT 1328 1328 V -> L (IN REF. 4).
FT CONFLICT 1573 1573 Y -> H (IN REF. 4).
FT CONFLICT 1623 1623
SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;

Query March 52.7%; Score 58; DB 1; Length 1707;
Best Local Similarity 61.1%; Pred. No. 0.49;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19

Db 1542 LYCNPQGVVYASRNDKS 1559

Search completed: April 5, 2004, 06:59:42
Job time : 4.39225 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 11.4092 Seconds
(without alignments)
525.440 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFCNVNVCNPNASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioplasmid:*
17: sp_archaeplastid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|---------------------|
| 1 | 93 | 84.5 | 212 | Q28512 | Q28512 macaca mula |
| 2 | 93 | 84.5 | 245 | Q9NYC4 | Q9NYC4 homo sapien |
| 3 | 92 | 83.6 | 161 | Q61430 | Q61430 mus musculus |
| 4 | 92 | 83.6 | 203 | Q29032 | Q29032 sus scrofa |
| 5 | 92 | 83.6 | 203 | Q28682 | Q28682 oryctolagus |
| 6 | 92 | 83.6 | 210 | Q28273 | Q28273 canis famil |
| 7 | 92 | 83.6 | 212 | Q28567 | Q28567 ovine aries |
| 8 | 92 | 83.6 | 225 | Q28271 | Q28271 canis famil |
| 9 | 92 | 83.6 | 226 | Q991Q8 | Q991Q8 mus musculus |
| 10 | 92 | 83.6 | 229 | Q8NF88 | Q8NF88 homo sapien |
| 11 | 92 | 83.6 | 229 | Q9NYC5 | Q9NYC5 homo sapien |
| 12 | 92 | 83.6 | 246 | Q61435 | Q61435 mus musculus |
| 13 | 92 | 83.6 | 979 | Q919K3 | Q919K3 gallus gall |
| 14 | 92 | 83.6 | 1075 | Q86X41 | Q86X41 homo sapien |
| 15 | 92 | 83.6 | 1621 | Q9H4R9 | Q9H4R9 homo sapien |
| 16 | 92 | 83.6 | 1669 | Q9QZS0 | Q9QZS0 mus musculus |

| | | | | | | |
|----|----|------|------|----|--------|---------------------|
| 17 | 90 | 81.8 | 179 | 11 | P70165 | P70165 mus musculus |
| 18 | 90 | 81.8 | 253 | 11 | Q61436 | Q61436 mus musculus |
| 19 | 90 | 81.8 | 585 | 11 | Q80V57 | Q80V57 mus musculus |
| 20 | 90 | 81.8 | 799 | 11 | Q8ENS7 | Q8ENS7 mus musculus |
| 21 | 90 | 81.8 | 886 | 4 | Q9NUB7 | Q9NUB7 homo sapien |
| 22 | 90 | 81.8 | 1684 | 6 | Q8HYC1 | Q8HYC1 canis famil |
| 23 | 90 | 81.8 | 1688 | 6 | Q866Z2 | Q866Z2 canis famil |
| 24 | 90 | 81.8 | 1691 | 11 | Q9ESQ2 | Q9ESQ2 mus musculus |
| 25 | 88 | 80.0 | 220 | 11 | Q63122 | Q63122 rattus norv |
| 26 | 80 | 72.7 | 1752 | 5 | Q07265 | Q07265 strongyloce |
| 27 | 76 | 69.1 | 1747 | 5 | Q26640 | Q26640 strongyloce |
| 28 | 69 | 62.7 | 1802 | 5 | Q17163 | Q17163 brugia mala |
| 29 | 64 | 58.2 | 205 | 6 | Q28274 | Q28274 canis famil |
| 30 | 64 | 58.2 | 546 | 11 | Q99K97 | Q99K97 mus musculus |
| 31 | 64 | 58.2 | 1600 | 4 | Q9UEH6 | Q9UEH6 homo sapien |
| 32 | 64 | 58.2 | 1691 | 11 | Q9ESQ1 | Q9ESQ1 mus musculus |
| 33 | 59 | 53.6 | 332 | 11 | Q64457 | Q64457 mus musculus |
| 34 | 59 | 53.6 | 1024 | 5 | Q8T7S4 | Q8T7S4 anopheles g |
| 35 | 59 | 53.6 | 1682 | 11 | Q9QZR9 | Q9QZR9 mus musculus |
| 36 | 59 | 53.6 | 1723 | 5 | Q9GQB1 | Q9GQB1 hydra atten |
| 37 | 59 | 53.6 | 1779 | 5 | Q9VMV4 | Q9VMV4 drosophila |
| 38 | 58 | 52.7 | 202 | 6 | Q28272 | Q28272 canis famil |
| 39 | 58 | 52.7 | 208 | 6 | Q29468 | Q29468 canis famil |
| 40 | 58 | 52.7 | 358 | 11 | Q91V13 | Q91V13 mus musculus |
| 41 | 58 | 52.7 | 673 | 4 | Q14052 | Q14052 homo sapien |
| 42 | 54 | 49.1 | 713 | 5 | Q9GV24 | Q9GV24 sarcophaga |
| 43 | 53 | 48.2 | 310 | 11 | Q8JZZ6 | Q8JZZ6 mus musculus |
| 44 | 51 | 46.4 | 2275 | 3 | Q93937 | Q93937 emericella |
| 45 | 49 | 44.5 | 346 | 4 | Q9NS55 | Q9NS55 homo sapien |

ALIGNMENTS

RESULT 1

ID Q28512 PRELIMINARY; PRT; 212 AA.
AC Q28512;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Fusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different mammals";
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
EL EMBL; L47280; AAA93861.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagen4 C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; Procollagen4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PROSITE00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357B64 CRC64;

Query Match 84.5%; Score 93; DB 6; Length 212;
Best Local Similarity 94.4%; Pred. No. 1.7e-07;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
||||| |||||||
Db 45 LFCNVNVCNCFASRNDYS 62

RESULT 2

Q9NYC4 PRELIMINARY; PRT; 245 AA.
ID Q9NYC4
AC Q9NYC4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tuncstatin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,
RA Erickson M.D., Hoffer H., Xiao Y., Stillman I.E., Kalluri R.,
RT "Distinct anti-tumor properties of a type IV collagen domain derived
RT from basement membrane.";
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL; AR258351; AAF72632.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 84.5%; Score 93; DB 4; Length 245;
Best Local Similarity 94.4%; Pred.No. 1.9e-07;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
||||| |||||||
Db 78 LFCNVNVCNCFASRNDYS 95

RESULT 3

Q61430 PRELIMINARY; PRT; 161 AA.
ID Q61430
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Oberbaumer I.;
RA STRAIN=129;
RT "Cloning of the NC1 domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82205; CAA57689.1; -;
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFB9236C5 CRC64;

Query Match 83.6%; Score 92; DB 11; Length 161;
Best Local Similarity 88.9%; Pred.No. 1.9e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
||||| |||||||
Db 12 LFCNVNVCNCFASRNDYS 29

RESULT 4

Q29032 PRELIMINARY; PRT; 203 AA.
ID Q29032
AC Q29032;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47284; AAA91882.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 83.6%; Score 92; DB 6; Length 203;
Best Local Similarity 88.9%; Pred.No. 2.3e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
||||| |||||||
Db 45 LFCNVNVCNCFASRNDYS 62

RESULT 5

Q28682 PRELIMINARY; PRT; 203 AA.
ID Q28682
AC Q28682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.

```

OS Oryctolagus cuniculus (Rabbit);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47283; AAA91893.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 83.6%; Score 92; DB 6; Length 203;
Best Local Similarity 88.9%; Pred. No. 2.3e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDS 19
DB 45 LFCNINDVCNFCASRNDS 62

RESULT 6
ID Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119B4CA823633D CRC64;

Query Match 83.6%; Score 92; DB 6; Length 210;

```

```

Best Local Similarity 88.9%; Pred. No. 2.4e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDS 19
DB 55 LFCNINNVCNFCASRNDS 72

RESULT 7
ID Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 83.6%; Score 92; DB 6; Length 212;
Best Local Similarity 88.9%; Pred. No. 2.4e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDS 19
DB 45 LFCNINDVCNFCASRNDS 62

RESULT 8
ID Q28271 PRELIMINARY; PRT; 225 AA.
AC Q28271;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains

```


RT of collagen type IV. Evidence from a canine model of X-linked
 RT nephritis with a COL4A5 gene mutation."
 RL J. Biol. Chem. 271:13821-13828(1996).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-Samoyed,
 RA Thorner P.S.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U50933; AAC48583.2; -.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro: IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD003923; ProcollagnC4; 2.
 DR SMART; SM00111; C4; 2.
 DR Collagen.
 KW Collagen.
 FT NON_TER 1
 FT NON_TER 225
 SQ SEQUENCE 225 AA; 24585 MW; 2C20455890416E47 CRC64;

Query Match 83.6%; Score 92; DB 6; Length 225;
 Best Local Similarity 88.9%; Pred. No. 2.6e-07;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNNVNCVCFASRNDYS 19
 |||||:|||||
 Db 69 LFCNNVNCVCFASRNDYS 86

RESULT 9

ID Q99LQ8 PRELIMINARY; PRT; 226 AA.
 AC Q99LQ8;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN COL4A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC02269; AA02269.1; -.
 DR MGD; MGI:88454; Col4a1.
 DR GO; GO:0005604; C:basement membrane; IEA.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 DR Hypothetical protein.
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 226
 SQ SEQUENCE 226 AA; 25042 MW; 4F7F0D5371181C21 CRC64;

Query Match 83.6%; Score 92; DB 11; Length 226;
 Best Local Similarity 88.9%; Pred. No. 2.6e-07;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNNVNCVCFASRNDYS 19
 |||||:|||||
 Db 60 LFCNNVNCVCFASRNDYS 77

RESULT 10

ID Q8NF88 PRELIMINARY; PRT; 229 AA.
 AC Q8NF88;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Arresten (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA He A.B.;
 RT "Cloning and Expression of Arresten in Escherichia coli and Pachia
 RT pastoris."
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF536207; AM97359.1; -.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 DR NON_TER 1
 FT NON_TER 229
 SQ SEQUENCE 229 AA; 25391 MW; 09B21FD5AB517E9E CRC64;

Query Match 83.6%; Score 92; DB 4; Length 229;
 Best Local Similarity 88.9%; Pred. No. 2.6e-07;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNNVNCVCFASRNDYS 19
 |||||:|||||
 Db 63 LFCNNVNCVCFASRNDYS 80

RESULT 11

ID Q9NYC5 PRELIMINARY; PRT; 229 AA.
 AC Q9NYC5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arresten (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H.,
 RA Takahashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K.,
 RA Erickson M.B., Dhanabal M., Simons M., Post M., Kufe D.,
 RA Weichselbaum R.R., Sukhatme V.P., Kalluri R.;
 RT "Anti-angiogenic cues from vascular basement membrane collagen."
 RL Cancer Res. 0:0-0(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Fu J., Bai X., Wang W., Ruan C.;
 RT "Arresten, a collagen-derived inhibitor of angiogenesis."
 RL Chung Hua Hsueh Yen Hsueh Tea Chih 22:0-0(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Peng X., Yin B., Yuan J., Qiang B.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Zheng Q.C., Song Z.F., Zheng Y.W., Li Y.Q., Shu X.;
 RT "Molecular cloning and sequencing of human arresten gene."
 RL Zhonghua Shi Yan Wei Ke Za Zhi 19:46-47(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Song Z.F., Zheng Q.C.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF258349; AAF72830.1; -.
 DR EMBL; AF363672; AAK53382.1; -.
 DR EMBL; AF400431; AAK92480.1; -.
 DR EMBL; AY285780; AAP43112.1; -.
 DR GO; GO:0005581; C:collagen; IEA.

```
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5C1D5 CRC64;

Query Match      83.6%; Score 92; DB 4; Length 229;
Best Local Similarity 88.9%; Pred. No. 2.6e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LFCNVNVCNFCASRNDYS 19
DB      |||||:|||||
        53 LFCNINNVNFCASRNDYS 80

RESULT 12
Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H.; Sanes J.R.; and alpha 5 chains in rodent basal
RT laminae: Sequence, distribution, association with laminins, and
RT developmental switches."
RL J. Cell Biol. 127:879-891(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z35166; CAAB4529.1; -.
DR PIR; I48302; I48302.
DR MGD; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON TER 1
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F58367324 CRC64;

Query Match      83.6%; Score 92; DB 11; Length 246;
Best Local Similarity 88.9%; Pred. No. 2.8e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LFCNVNVCNFCASRNDYS 19
DB      |||||:|||||
        79 LFCNINNVNFCASRNDYS 96

RESULT 13
Q919K3 PRELIMINARY; PRT; 979 AA.
AC Q919K3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE Collagen IV al chain (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M.; Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
RT basal lamina."
RL Dev. Biol. 0:0-0(2000).
DR EMBL; AF239838; AAF44681.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Procollagn4_C.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; C1g_helix; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match      83.6%; Score 92; DB 13; Length 979;
Best Local Similarity 88.9%; Pred. No. 9.8e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LFCNVNVCNFCASRNDYS 19
DB      |||||:|||||
        813 LFCNINNVNFCASRNDYS 830

RESULT 14
Q86X41 PRELIMINARY; PRT; 1075 AA.
AC Q86X41;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047305; AAH47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Collagen; 13.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match      83.6%; Score 92; DB 4; Length 1075;
Best Local Similarity 88.9%; Pred. No. 1.1e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 LFCNVNVCNFCASRNDYS 19
```

```

Db      909 LFCNINVCNFSRNDYS 926
||||:| |||||
RESULT 15
Q9H4R9 PRELIMINARY; PRT; 1621 AA.
AC Q9H4R9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE BA472K17.2 (Collagen type IV alpha 1) (Fragment).
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390755; CAC13153.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g_helix; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1621 AA; 155705 MW; 73F6F901CDOEDBA2 CRC64;
Query Match 83.6%; Score 92; DB 4; Length 1621;
Best Local Similarity 88.9%; Pred. No. 1.6e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 2 LFCNVCNCFASRNDYS 19
||||:| |||||
Db      1455 LFCNINVCNFSRNDYS 1472

```

Search completed: April 5, 2004, 07:03:58
 Job time : 11.4092 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 17.1138 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFCNVNVCNCFASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 110 | 100.0 | 19 | 6 | ADA20240 TP3 peptid |
| 2 | 93 | 84.5 | 25 | 6 | ADA20236 |
| 3 | 93 | 84.5 | 27 | 6 | ADA20238 |
| 4 | 93 | 84.5 | 79 | 5 | Aau75600 Human typ |
| 5 | 93 | 84.5 | 79 | 6 | ADA20264 Human typ |
| 6 | 93 | 84.5 | 88 | 5 | Aau75608 Human typ |
| 7 | 93 | 84.5 | 88 | 5 | Aau75607 Human typ |
| 8 | 93 | 84.5 | 88 | 6 | ADA20271 Human typ |
| 9 | 93 | 84.5 | 88 | 6 | ADA20272 Human typ |
| 10 | 93 | 84.5 | 124 | 5 | Aau75594 Human typ |
| 11 | 93 | 84.5 | 124 | 6 | ADA20258 Human typ |
| 12 | 93 | 84.5 | 132 | 5 | Aau75597 Human typ |
| 13 | 93 | 84.5 | 132 | 6 | ADA20261 Human typ |
| 14 | 93 | 84.5 | 131 | 5 | Aau75596 Human typ |
| 15 | 93 | 84.5 | 191 | 6 | ADA20260 Human typ |
| 16 | 93 | 84.5 | 211 | 3 | AAY95918 Human Goo |
| 17 | 93 | 84.5 | 211 | 5 | ABG79208 Human GP |
| 18 | 93 | 84.5 | 218 | 2 | AAR79164 Partial s |
| 19 | 93 | 84.5 | 218 | 2 | AAY44172 Human typ |
| 20 | 93 | 84.5 | 218 | 3 | AAY56784 Human alp |
| 21 | 93 | 84.5 | 218 | 4 | Aae09484 Human alp |
| 22 | 93 | 84.5 | 232 | 7 | ADC17697 Human typ |
| 23 | 93 | 84.5 | 244 | 5 | ABG79218 Human typ |
| 24 | 93 | 84.5 | 244 | 5 | ABG79219 Human Goo |
| 25 | 93 | 84.5 | 244 | 5 | ABG79217 Human typ |

| | | | | | |
|----|----|------|------|---|--------------------|
| 26 | 93 | 84.5 | 244 | 5 | Aau75595 Human typ |
| 27 | 93 | 84.5 | 244 | 6 | ADA20235 Human typ |
| 28 | 93 | 84.5 | 245 | 3 | AAY67942 Human typ |
| 29 | 93 | 84.5 | 245 | 5 | Aau75589 Human typ |
| 30 | 93 | 84.5 | 254 | 5 | Aau75598 Human typ |
| 31 | 93 | 84.5 | 268 | 2 | AAY31993 Type IV c |
| 32 | 93 | 84.5 | 268 | 3 | AAY97555 Human alp |
| 33 | 93 | 84.5 | 1670 | 7 | ADA47063 Human Pro |
| 34 | 92 | 83.6 | 229 | 1 | AAP33524 Complete |
| 35 | 92 | 83.6 | 229 | 3 | AAY67943 Human typ |
| 36 | 92 | 83.6 | 229 | 5 | Aau75587 Human typ |
| 37 | 92 | 83.6 | 229 | 6 | ADA20217 Human typ |
| 38 | 92 | 83.6 | 229 | 7 | ADC17695 Human typ |
| 39 | 92 | 83.6 | 260 | 2 | AAY31991 Type IV c |
| 40 | 92 | 83.6 | 260 | 3 | AAY97553 Human alp |
| 41 | 92 | 83.6 | 406 | 3 | AAB58169 Lung canc |
| 42 | 92 | 83.6 | 471 | 2 | AAR79163 Partial s |
| 43 | 92 | 83.6 | 471 | 2 | AAY44171 Bovine ty |
| 44 | 92 | 83.6 | 471 | 3 | AAY56783 Bovine al |
| 45 | 92 | 83.6 | 471 | 4 | Aae09483 Bovine al |

ALIGNMENTS

RESULT 1

ADA20240

ID ADA20240 standard; peptide; 19 AA.

XX AC ADA20240;

XX DT 20-NOV-2003 (first entry)

XX DE TP3 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

KW metastasis; basement membrane organisation; type IV collagen network;

KW C-terminal globular non-collagenous domain; NCI; type IV collagen;

KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

KW cytostatic; gene therapy; TP3 peptide; tumstatin; human;

KW type IV collagen alpha 3 chain; mutant; mutein.

XX OS Synthetic.

XX OS Homo sapiens.

XX PH Key

XX Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Phe substituted by Lys"

FT Misc-difference 8 /note= "Wild-type Asp substituted by Cys"

FT WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX XX WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor

growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 64; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the T7 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.

XX SQ Sequence 19 AA;

Query Match 100.0%; Score 110; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLFNVNVCNCFASRNDYS 19
 |||||
 DB 1 KLFNVNVCNCFASRNDYS 19

RESULT 2

ADA20236
 ID ADA20236 standard; peptide; 25 AA.

XX AC ADA20236;

DT 20-NOV-2003 (first entry)

DE T7 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; T7 peptide; tumstatin; human;
 KW type IV collagen alpha 3 chain.

OS Homo sapiens.

XX WO2003059257-A2.

PN 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 53; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the T7 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.

XX SQ Sequence 25 AA;

Query Match 84.5%; Score 93; DB 6; Length 25;
 Best Local Similarity 94.4%; Pred. No. 5.8e-07;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19
 |||||
 DB 5 LFCNVNVCNCFASRNDYS 22

RESULT 3

ADA20238
 ID ADA20238 standard; peptide; 27 AA.

XX AC ADA20238;

DT 20-NOV-2003 (first entry)

DE T8 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; T8 peptide; tumstatin; human;
 KW type IV collagen alpha 3 chain; mutant; mutein.

OS Synthetic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 62; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC from pre-existing vessels is essential for tumour growth and metastasis.


```

XX
PI Kalluri R;
XX
XX
DR WPI; 2003-587256/55.
DR N-PSDB; ADA20224.
XX
XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Claim 94; SEQ ID NO 26; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tum-5, an abridged form of the "tumstatin" protein of
XX the invention which was derived from the amino acid sequence of the alpha
XX 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does
XX not appear in the specification but was created by the indexer from
XX information given in the specification.
XX
SQ Sequence 79 AA;
    Query Match      84.5%; Score 93; DB 6; Length 79;
    Best Local Similarity 94.4%; Pred. No. 1.9e-06;
    Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19
    ||||| ||||| ||||| |||||
DB 24 LFCNVNVCNCFASRNDYS 41

RESULT 6
AAU75608
ID AAU75608 standard; protein; 88 AA.
XX
AC AAU75608;
XX
XX 08-MAY-2002 (first entry)
XX
XX Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.
XX
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 82 /note="Wild type Cys substituted with Ala"
XX
XX WO200151523-A2.
XX
XX 19-JUL-2001.
XX
XX 08-JAN-2001; 2001WO-US0000565.
XX
XX 07-JAN-2000; 2000US-00479118.
XX
XX 04-APR-2000; 2000US-00543371.
XX

```

```

PR 21-JUL-2000; 2000US-00625191.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2002-189037/24.
XX
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX
XX Claim 41; Page 153; 205pp; English.
XX
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell apoptosis in a tissue, where the
XX promoting or inducing endothelial cell apoptosis by one or more endothelial cell
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in a vertebrate,
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX or allelic variants are useful in the preparation of a medicament for
XX treating a disorder involving inhibiting angiogenesis in a tissue, where
XX the angiogenesis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits; or by promoting or
XX inducing endothelial cell apoptosis in a tissue, where the endothelial
XX cell apoptosis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits. The medicament is useful
XX in inhibiting tumour growth and for the regression of an established
XX tumour. The present sequence represents the amino acid sequence of human
XX type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
XX consists of residues 5-126 of Tumstatin
XX
SQ Sequence 88 AA;
    Query Match      84.5%; Score 93; DB 5; Length 88;
    Best Local Similarity 94.4%; Pred. No. 2.1e-06;
    Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19
    ||||| ||||| ||||| |||||
DB 34 LFCNVNVCNCFASRNDYS 51

RESULT 7
AAU75607
ID AAU75607 standard; protein; 88 AA.
XX
XX AAU75607;
XX
XX 08-MAY-2002 (first entry)
XX
XX Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.
XX

```

KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphabeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
OS Homo sapiens.
XX WO200151523-A2.
XX 19-JUL-2001.
XX 08-JAN-2001; 2001WO-US000565.
XX 07-JAN-2000; 2000US-00479118.
XX 04-APR-2000; 2000US-00543371.
XX 21-JUL-2000; 2000US-00625191.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2002-188037/24.
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX Claim 32; Page 152; 205pp; English.
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphabeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; or (b) by
XX promoting or inducing endothelial cell apoptosis in a tissue, where the
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alphav, betav or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in the preparation
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX or allelic variants are useful in the preparation of a medicament for
XX treating a disorder involving inhibiting angiogenesis in a tissue, where
XX the angiogenesis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits; or by promoting or
XX inducing endothelial cell apoptosis in a tissue, where the endothelial
XX cell apoptosis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits. The medicament is useful
XX in inhibiting tumour growth and for the regression of an established
XX tumour. The present sequence represents the amino acid sequence of human
XX type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists
XX of residues 45-132 of Tumstatin
SQ Sequence 88 AA;

Query Match 84.5%; Score 93; DB 5; Length 88;
Best Local Similarity 94.4%; Pred. No. 2.1e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 34 LFCNVNVCNCFASRNDYS 51
RESULT 8
ADA20271
ID ADA20271 standard; protein; 88 AA.
XX ADA20271;
AC ADA20271;
XX 20-NOV-2003 (first entry)
XX Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human;
XX tumstatin 45-132.
XX Homo sapiens.
XX OS
XX PN WO2003059257-A2.
XX 24-JUL-2003.
XX 20-DEC-2002; 2002WO-US040938.
XX 21-DEC-2001; 2001US-00032221.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2003-587256/55.
XX N-PSDB; ADA20224.
XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX Claim 94; SEQ ID NO 33; 240pp; English.
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"
XX protein of the invention which was derived from the amino acid sequence
XX of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
XX ID33) does not appear in the specification but was created by the indexer
XX from information given in the specification.
XX SQ Sequence 88 AA;

Query Match 84.5%; Score 93; DB 6; Length 88;
Best Local Similarity 94.4%; Pred. No. 2.1e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19

Db 33 LFCNVNVCNFCASRNDYS 50

||||| ||||| ||||| ||||| |||||

RESULT 9
ADA20272
ID ADA20272 standard; protein; 88 AA.
XX
AC ADA20272;
XX
XX 20-NOV-2003 (first entry)
XX
XX Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.
XX
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human;
KW tumstatin 5-125-C-A; mutant; mutein.
XX
XX Synthetic.
OS
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 81
FT /notes "wild-type Cys substituted by Ala at position 125
FT of full-length tumstatin"
XX
XX W02003059257-A2.
XX
XX 24-JUL-2003.
XX
XX 20-DEC-2002; 2002W0-US040938.
XX
XX 21-DEC-2001; 2001US-00032221.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2003-587256/55.
XX
XX New peptide, useful for preparing a composition for inhibiting tumor
FT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
XX Claim 94; SEQ ID NO 34; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
CC the "tumstatin" protein of the invention which was derived from the amino
CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This
CC sequence (Seq ID33) does not appear in the specification but was created
CC by the indexer from information given in the specification.
XX
XX Sequence 88 AA;
SQ

Best Local Similarity 94.4%; Pred. No. 2.1e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19
||||| ||||| ||||| ||||| |||||
Db 33 LFCNVNVCNFCASRNDYS 50

RESULT 10
AAU75594
ID AAU75594 standard; protein; 124 AA.
XX
AC AAU75594;
XX
XX 08-MAY-2002 (first entry)
XX
XX Human type IV collagen alpha 3 chain mutant, Tumstatin 333.
DE
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
XX Homo sapiens.
OS
XX W0200151523-A2.
XX
XX 19-JUL-2001.
XX
XX 08-JAN-2001; 2001W0-US000565.
XX
XX 07-JAN-2000; 2000US-00479118.
PR 04-APR-2000; 2000US-00543371.
PR 21-JUL-2000; 2000US-00625191.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2002-188037/24.
XX
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
FT treating disorders involving angiogenesis.
XX
XX Example 33; Page; 205pp; English.
XX
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

Query Match 84.5%; Score 93; DB 6; Length 88;

CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of
 CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in
 CC the specification but is derived from the wild type human Tumstatin
 CC sequence given in figure 18A (see AAU75589)
 XX
 SQ Sequence 124 AA;

Query Match 84.5%; Score 93; DB 5; Length 124;
 Best Local Similarity 94.4%; Pred. No. 2.9e-06;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNVNVCNCFASRNDYS 19
 ||||| ||||| ||||| |||||
 Db 77 LFCNVNVCNCFASRNDYS 94

RESULT 11
 ADA20258
 ID ADA20258 standard; protein; 124 AA.
 AC ADA20258;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.
 XX
 OS Homo sapiens.
 XX
 PN WO2003059257-A2.
 XX
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PR 21-DEC-2001; 2001US-00032221.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX
 PI Kalluri R;
 XX
 DR WPI; 2003-587256/55.
 DR N-PSDB; ADA20224.
 XX
 PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 PS
 PS Claim 94; SEQ ID NO 20; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collageneous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"
 CC protein of the invention which was derived from the amino acid sequence
 CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
 CC ID20) does not appear in the specification but was created by the indexer
 CC from information given in the specification.
 XX
 SQ Sequence 124 AA;

Query Match 84.5%; Score 93; DB 6; Length 124;
 Best Local Similarity 94.4%; Pred. No. 2.9e-06;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LFCNVNVCNCFASRNDYS 19
 ||||| ||||| ||||| |||||
 Db 77 LFCNVNVCNCFASRNDYS 94

RESULT 12
 AAU75597
 ID AAU75597 standard; protein; 132 AA.
 AC AAU75597;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain mutant, Tum-2.
 XX
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.
 OS
 OS Homo sapiens.
 XX
 PN WO200151523-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 08-JAN-2001; 2001WO-US000565.
 XX
 PR 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX
 PI Kalluri R;
 XX
 DR WPI; 2002-188037/24.
 XX
 PT A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
 PT treating disorders involving angiogenesis.
 PS
 PS Claim 31; Page 152; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arretsen, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arretsen, Canstatin or Tumstatin; (4) use of one
 CC or more soluble receptors that bind Arretsen, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arretsen, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 18A (see AAU75589)

XX SQ Sequence 132 AA;

Query Match 84.5%; Score 93; DB 5; Length 132;

Best Local Similarity 94.4%; Pred. No. 3.1e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19

||||| |||||||

Db 78 LFCNVNVCNCFASRNDYS 95

RESULT 13

ADA20261

ID ADA20261 standard; protein; 132 AA.

XX AC ADA20261;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX metastasis; basement membrane organisation; type IV collagen network;

XX C-terminal globular non-collagenous domain; NCI; type IV collagen;

XX cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX OS Homo sapiens.

XX PN WQ2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.

XX SQ Sequence 132 AA;

Query Match 84.5%; Score 93; DB 6; Length 132;

Best Local Similarity 94.4%; Pred. No. 3.1e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19

||||| |||||||

Db 77 LFCNVNVCNCFASRNDYS 94

RESULT 14

AAU75596

ID AAU75596 standard; protein; 191 AA.

XX AC AAU75596;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

XX non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;

XX endothelial cell proliferation; apoptosis; Arretsen; Canstatin;

XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX OS Homo sapiens.

XX PN WQ200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
 PT treating disorders involving angiogenesis.

XX PS Example 32; Page; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause

CC apoptosis of endothelial cells. Also described are the following: (1) use

CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for

CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by

CC promoting or inducing endothelial cell apoptosis in a tissue, where the

CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of

CC an antibody or peptide that specifically binds the alpha1, alpha2,

CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the

CC preparation of a medicament for inhibiting angiogenesis or cell

CC proliferation; (3) use of an inhibitor, such as an antibody, antibody

CC fragment or peptide of receptor-mediated angiogenesis in the preparation

CC of a medicament for treating a proliferative disease in a vertebrate,

CC where the disease is characterised by angiogenesis that is mediated by

CC receptors to Arresten, Canstatin or Tumstatin and where the receptors

CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one

CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in

CC the presence of a medicament for promoting angiogenesis in a tissue; and

CC (5) use of integrins in the preparation of a medicament for promoting or

CC inducing angiogenesis or cell proliferation in a tissue. The fragments

CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

CC or allelic variants are useful in the preparation of a medicament for

CC treating a disorder involving inhibiting angiogenesis in a tissue, where

CC the angiogenesis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits; or by promoting or

CC inducing endothelial cell apoptosis in a tissue, where the endothelial

CC cell apoptosis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits. The medicament is useful

CC in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human

CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of

CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in

CC the specification but is derived from the wild type human Tumstatin

CC sequence given in figure 18A (see AAU75589)

XX SQ Sequence 191 AA;

Query Match 84.5%; Score 93; DB 5; Length 191;

Best Local Similarity 94.4%; Pred. No. 4.5e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19

Db 25 LFCNVNVCNCFASRNDYS 42

RESULT 15

ADA20260

ID ADA20260 standard; protein; 191 AA.

XX AC ADA20260;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-1 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX KW metastasis; basement membrane organisation; type IV collagen network;

XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;

XX KW tumstatin N53.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 22; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries

CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-

CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular

CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

CC collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cytostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present

CC sequence is that of tum-1 (tumstatin N53), an abridged form of the

CC "tumstatin" protein of the invention which was derived from the amino

CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This

CC sequence (Seq ID22) does not appear in the specification but was created

CC by the indexer from information given in the specification.

XX SQ Sequence 191 AA;

Query Match 84.5%; Score 93; DB 6; Length 191;

Best Local Similarity 94.4%; Pred. No. 4.5e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19

Db 24 LFCNVNVCNCFASRNDYS 41

Search completed: April 5, 2004, 06:58:32

Job time : 17.1138 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 11.9153 Seconds
(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFCNVNCVCFASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 110 | 100.0 | 19 | 14 | US-10-032-221B-41 |
| 2 | 93 | 84.5 | 25 | 14 | Sequence 37, Appl |
| 3 | 93 | 84.5 | 27 | 14 | Sequence 39, Appl |
| 4 | 93 | 84.5 | 79 | 14 | US-10-032-221B-39 |
| 5 | 93 | 84.5 | 88 | 14 | Sequence 26, Appl |
| 6 | 93 | 84.5 | 88 | 14 | US-10-032-221B-26 |
| 7 | 93 | 84.5 | 124 | 14 | US-10-032-221B-33 |
| 8 | 93 | 84.5 | 132 | 14 | Sequence 34, Appl |
| 9 | 93 | 84.5 | 191 | 14 | US-10-032-221B-20 |
| 10 | 93 | 84.5 | 211 | 14 | Sequence 22, Appl |
| 11 | 93 | 84.5 | 211 | 14 | US-10-032-221B-22 |
| 12 | 93 | 84.5 | 232 | 14 | Sequence 46, Appl |
| 13 | 93 | 84.5 | 244 | 14 | US-10-032-837-46 |
| 14 | 92 | 83.6 | 229 | 14 | US-10-032-221B-10 |
| 15 | 92 | 83.6 | 229 | 14 | Sequence 302, Appl |
| | | | | | Sequence 2, Appl |

| | | | | | | |
|----|----|------|------|----|---------------------|-------------------|
| 16 | 92 | 83.6 | 406 | 9 | US-09-925-302-507 | Sequence 507, App |
| 17 | 92 | 83.6 | 1669 | 15 | US-10-372-683-8 | Sequence 8, Appli |
| 18 | 90 | 81.8 | 25 | 14 | US-10-032-221B-38 | Sequence 38, Appl |
| 19 | 90 | 81.8 | 46 | 9 | US-09-864-761-48095 | Sequence 48095, A |
| 20 | 90 | 81.8 | 229 | 14 | US-10-206-699-306 | Sequence 306, App |
| 21 | 90 | 81.8 | 309 | 9 | US-09-925-297-496 | Sequence 496, App |
| 22 | 82 | 74.5 | 18 | 14 | US-10-206-699-260 | Sequence 260, App |
| 23 | 82 | 74.5 | 22 | 14 | US-10-206-699-266 | Sequence 266, App |
| 24 | 81 | 73.6 | 18 | 14 | US-10-206-699-259 | Sequence 259, App |
| 25 | 81 | 73.6 | 22 | 14 | US-10-206-699-285 | Sequence 285, App |
| 26 | 79 | 71.8 | 18 | 14 | US-10-206-699-261 | Sequence 261, App |
| 27 | 79 | 71.8 | 22 | 14 | US-10-206-699-267 | Sequence 267, App |
| 28 | 75 | 68.2 | 1744 | 15 | US-10-369-493-5832 | Sequence 5832, Ap |
| 29 | 73 | 66.4 | 27 | 14 | US-10-032-221B-40 | Sequence 40, Appl |
| 30 | 71 | 64.5 | 1759 | 15 | US-10-369-493-7032 | Sequence 7032, Ap |
| 31 | 69 | 62.7 | 27 | 14 | US-10-032-221B-42 | Sequence 42, Appl |
| 32 | 64 | 58.2 | 142 | 9 | US-09-864-761-38021 | Sequence 38021, A |
| 33 | 64 | 58.2 | 228 | 14 | US-10-206-699-307 | Sequence 307, App |
| 34 | 62 | 56.4 | 18 | 14 | US-10-206-699-264 | Sequence 264, App |
| 35 | 62 | 56.4 | 22 | 14 | US-10-206-699-270 | Sequence 270, App |
| 36 | 61 | 55.5 | 14 | 14 | US-10-206-699-3 | Sequence 3, Appli |
| 37 | 61 | 55.5 | 15 | 14 | US-10-206-699-210 | Sequence 210, App |
| 38 | 61 | 55.5 | 15 | 14 | US-10-206-699-212 | Sequence 212, App |
| 39 | 61 | 55.5 | 18 | 14 | US-10-206-699-254 | Sequence 254, App |
| 40 | 61 | 55.5 | 20 | 14 | US-10-032-221B-30 | Sequence 30, Appl |
| 41 | 60 | 54.5 | 14 | 14 | US-10-206-699-2 | Sequence 2, Appli |
| 42 | 60 | 54.5 | 18 | 14 | US-10-206-699-253 | Sequence 253, App |
| 43 | 59 | 53.6 | 231 | 14 | US-10-206-699-305 | Sequence 305, App |
| 44 | 58 | 52.7 | 14 | 14 | US-10-206-699-4 | Sequence 4, Appli |
| 45 | 58 | 52.7 | 18 | 14 | US-10-206-699-255 | Sequence 255, App |

ALIGNMENTS

RESULT 1

US-10-032-221B-41
; Sequence 41, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: TP3 (amino acids 76-94 of SEQ ID NO:10; lysine has been substit
; OTHER INFORMATION: ed for the phenylalanine residue at position 76 of the full-len
; OTHER INFORMATION: h Tumstatin molecule, and cysteine has been substituted for th
; OTHER INFORMATION: aspartic acid at position 83)
US-10-032-221B-41

Query Match

100.0%; Score 110; DB 14; Length 19;

RESULT 3
US-10-032-221B-39
Sequence 39, Application US/10032221B
Publication No. US2003014481A1
GENERAL INFORMATION:
APPLICANT: Kalluri, Raghuram
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/335,224
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)
US-10-032-221B-33

Query Match 84.5%; Score 93; DB 14; Length 88;
Best Local Similarity 94.4%; Pred. No. 5.9e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 33 LFCNVNVCNCFASRNDYS 50

RESULT 6
US-10-032-221B-34
; Sequence 34, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine has been substituted for the cysteine residue at position 125 of the full-length Tumstatin molecule)
US-10-032-221B-34

Query Match 84.5%; Score 93; DB 14; Length 88;

Best Local Similarity 94.4%; Pred. No. 5.9e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 LFCNVNVCNCFASRNDYS 19
Db 33 LFCNVNVCNCFASRNDYS 50

RESULT 7
US-10-032-221B-20
; Sequence 20, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match 84.5%; Score 93; DB 14; Length 124;
Best Local Similarity 94.4%; Pred. No. 8.2e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 77 LFCNVNVCNCFASRNDYS 94

RESULT 8
US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25

; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match 84.5%; Score 93; DB 14; Length 132;
Best Local Similarity 94.4%; Pred. No. 8.7e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDS 19
Db 77 LFCNVNVCNCFASRNDS 94

RESULT 9

US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US2003014481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match 84.5%; Score 93; DB 14; Length 191;
Best Local Similarity 94.4%; Pred. No. 1.2e-05;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDS 19
Db 24 LFCNVNVCNCFASRNDS 41

RESULT 10

US-10-270-877-46
; Sequence 46, Application US/10270877
; Publication No. US2003004979A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877

; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match 84.5%; Score 93; DB 14; Length 211;
Best Local Similarity 94.4%; Pred. No. 1.3e-05;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDS 19
Db 77 LFCNVNVCNCFASRNDS 94

RESULT 11

US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US2003005448A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match 84.5%; Score 93; DB 14; Length 211;
Best Local Similarity 94.4%; Pred. No. 1.3e-05;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDS 19
Db 77 LFCNVNVCNCFASRNDS 94

RESULT 12

US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854

;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 304
;; LENGTH: 232
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; OTHER INFORMATION: alpha 3 chain
US-10-206-699-304

Query Match 84.5%; Score 93; DB 14; Length 232;
Best Local Similarity 94.4%; Pred. No. 1.5e-05;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19
||||| |||||||
Db 65 LFCNVNVCNFCASRNDYS 82
||||| |||||||

RESULT 13
US-10-032-221B-10
;; Sequence 10, Application US/10032221B
;; Publication No. US20030144481A1
;; GENERAL INFORMATION:
;; APPLICANT: Kalluri, Raghuram
;; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
;; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
;; CURRENT APPLICATION NUMBER: US/10/032,221B
;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: PCT/US01/00565
;; PRIOR FILING DATE: 2001-01-08
;; PRIOR APPLICATION NUMBER: US 09/625,191
;; PRIOR FILING DATE: 2000-07-21
;; PRIOR APPLICATION NUMBER: US 09/543,371
;; PRIOR FILING DATE: 2000-04-04
;; PRIOR APPLICATION NUMBER: US 09/479,118
;; PRIOR FILING DATE: 2000-01-07
;; PRIOR APPLICATION NUMBER: US 09/335,224
;; PRIOR FILING DATE: 1999-06-17
;; PRIOR APPLICATION NUMBER: US 60/126,175
;; PRIOR FILING DATE: 1999-03-25
;; PRIOR APPLICATION NUMBER: US 60/089,689
;; PRIOR FILING DATE: 1998-06-17
;; NUMBER OF SEQ ID NOS: 58
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 10
;; LENGTH: 244
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-032-221B-10

Query Match 84.5%; Score 93; DB 14; Length 244;
Best Local Similarity 94.4%; Pred. No. 1.5e-05;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19
||||| |||||||
Db 77 LFCNVNVCNFCASRNDYS 94
||||| |||||||

RESULT 14
US-10-206-699-302
;; Sequence 302, Application US/10206699
;; Publication No. US20030100510A1
;; GENERAL INFORMATION:
;; APPLICANT: Sundaramoorthy, M.
;; APPLICANT: Hudson, B.
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
;; FILE REFERENCE: MBHB 01-1017

;; CURRENT APPLICATION NUMBER: US/10/206,699
;; CURRENT FILING DATE: 2002-07-26
;; PRIOR APPLICATION NUMBER: US 60/308,523
;; PRIOR FILING DATE: 2001-07-27
;; PRIOR APPLICATION NUMBER: US 60/351,289
;; PRIOR FILING DATE: 2001-10-29
;; PRIOR APPLICATION NUMBER: US 60/366,854
;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 302
;; LENGTH: 229
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; OTHER INFORMATION: alpha 1 chain
US-10-206-699-302

Query Match 83.6%; Score 92; DB 14; Length 229;
Best Local Similarity 88.9%; Pred. No. 2e-05;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19
||||| |||||||
Db 63 LFCNVNVCNFCASRNDYS 80
||||| |||||||

RESULT 15
US-10-032-221B-2
;; Sequence 2, Application US/10032221B
;; Publication No. US20030144481A1
;; GENERAL INFORMATION:
;; APPLICANT: Kalluri, Raghuram
;; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
;; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
;; CURRENT APPLICATION NUMBER: US/10/032,221B
;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: PCT/US01/00565
;; PRIOR FILING DATE: 2001-01-08
;; PRIOR APPLICATION NUMBER: US 09/625,191
;; PRIOR FILING DATE: 2000-07-21
;; PRIOR APPLICATION NUMBER: US 09/543,371
;; PRIOR FILING DATE: 2000-04-04
;; PRIOR APPLICATION NUMBER: US 09/479,118
;; PRIOR FILING DATE: 2000-01-07
;; PRIOR APPLICATION NUMBER: US 09/335,224
;; PRIOR FILING DATE: 1999-06-17
;; PRIOR APPLICATION NUMBER: US 60/126,175
;; PRIOR FILING DATE: 1999-03-25
;; PRIOR APPLICATION NUMBER: US 60/089,689
;; PRIOR FILING DATE: 1998-06-17
;; NUMBER OF SEQ ID NOS: 58
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2
;; LENGTH: 229
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-032-221B-2

Query Match 83.6%; Score 92; DB 14; Length 229;
Best Local Similarity 88.9%; Pred. No. 2e-05;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19
||||| |||||||
Db 63 LFCNVNVCNFCASRNDYS 80
||||| |||||||

Search completed: April 5, 2004, 07:36:07
Job time : 11.9153 secs

us-10-032-221b-41.rapb

Mon Apr 5 07:53:15 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 4.46247 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFCNVNVCNPFASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 93 | 84.5 | 211 | 4 | US-09-512-563C-46 |
| 2 | 93 | 84.5 | 218 | 2 | US-08-399-889-25 |
| 3 | 93 | 84.5 | 218 | 3 | US-09-167-364-25 |
| 4 | 93 | 84.5 | 218 | 3 | US-09-439-897-4 |
| 5 | 93 | 84.5 | 268 | 4 | US-09-589-927-6 |
| 6 | 93 | 84.5 | 268 | 4 | US-09-277-665-6 |
| 7 | 93 | 84.5 | 268 | 4 | US-09-589-987-6 |
| 8 | 92 | 83.6 | 260 | 4 | US-09-589-927-2 |
| 9 | 92 | 83.6 | 260 | 4 | US-09-277-665-2 |
| 10 | 92 | 83.6 | 260 | 4 | US-09-589-987-2 |
| 11 | 92 | 83.6 | 471 | 2 | US-08-399-889-24 |
| 12 | 92 | 83.6 | 471 | 3 | US-09-167-364-24 |
| 13 | 92 | 83.6 | 471 | 3 | US-09-439-897-2 |
| 14 | 90 | 81.8 | 264 | 4 | US-09-589-927-10 |
| 15 | 90 | 81.8 | 264 | 4 | US-09-277-665-10 |
| 16 | 90 | 81.8 | 264 | 4 | US-09-589-987-10 |
| 17 | 84 | 58.2 | 260 | 4 | US-09-589-927-12 |
| 18 | 64 | 58.2 | 260 | 4 | US-09-277-665-12 |
| 19 | 64 | 58.2 | 260 | 4 | US-09-589-987-12 |
| 20 | 62 | 56.4 | 1694 | 1 | US-08-494-168-2 |
| 21 | 59 | 53.6 | 260 | 4 | US-09-589-927-8 |
| 22 | 59 | 53.6 | 260 | 4 | US-09-277-665-8 |
| 23 | 59 | 53.6 | 260 | 4 | US-09-589-987-8 |
| 24 | 58 | 52.7 | 258 | 4 | US-09-589-927-4 |
| 25 | 58 | 52.7 | 258 | 4 | US-09-277-665-4 |
| 26 | 58 | 52.7 | 258 | 4 | US-09-589-987-4 |
| 27 | 49 | 44.5 | 347 | 4 | US-09-636-215-590 |

```

28 49 44.5 347 4 US-09-685-166A-590 Sequence 590, Appl
29 49 44.5 374 2 US-08-820-170A-25 Sequence 25, Appl
30 49 44.5 374 3 US-09-055-899-25 Sequence 25, Appl
31 49 44.5 374 3 US-09-273-565-25 Sequence 25, Appl
32 49 44.5 374 4 US-09-565-538-25 Sequence 25, Appl
33 49 44.5 374 4 US-09-661-468-25 Sequence 25, Appl
34 49 44.5 374 4 US-09-976-165-25 Sequence 2, Appl
35 49 44.5 374 4 US-09-227-853A-2 Sequence 2, Appl
36 49 44.5 374 5 PCT-US95-06385-2 Sequence 2, Appl
37 45 40.9 69 4 US-09-621-976-5669 Sequence 5669, Ap
38 44 40.0 49 1 US-07-865-166A-6 Sequence 6, Appl
39 42 38.2 45 3 US-08-965-903B-19 Sequence 19, Appl
40 42 38.2 100 3 US-08-965-903B-11 Sequence 11, Appl
41 42 38.2 100 4 US-09-370-398-7 Sequence 7, Appl
42 42 38.2 100 4 US-10-090-190-7 Sequence 4, Appl
43 42 38.2 124 3 US-08-965-903B-4 Sequence 4, Appl
44 42 38.2 182 4 US-09-543-681A-6443 Sequence 6443, Ap
45 42 38.2 461 1 US-08-385-229-2 Sequence 2, Appl

```

ALIGNMENTS

RESULT 1
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 84.5%; Score 93; DB 4; Length 211;
Best Local Similarity 94.4%; Pred. No. 1.9e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNPFASRNDYS 19
DB 77 LFCNVNVCNPFASRNDYS 94

RESULT 2
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT

```
; ORGANISM: Human
US-08-399-889-25

Query Match      84.5%; Score 93; DB 2; Length 218;
Best Local Similarity 94.4%; Pred. No. 2e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19
Db 51 LFCNVNVCNCFASRDY 58

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match      84.5%; Score 93; DB 3; Length 218;
Best Local Similarity 94.4%; Pred. No. 2e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19
Db 51 LFCNVNVCNCFASRDY 58

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      84.5%; Score 93; DB 3; Length 218;
Best Local Similarity 94.4%; Pred. No. 2e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19
Db 51 LFCNVNVCNCFASRDY 58

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match      84.5%; Score 93; DB 4; Length 268;
Best Local Similarity 94.4%; Pred. No. 2.5e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19
Db 101 LFCNVNVCNCFASRDY 118

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match      84.5%; Score 93; DB 4; Length 268;
Best Local Similarity 94.4%; Pred. No. 2.5e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19
Db 101 LFCNVNVCNCFASRDY 118

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6438140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match      84.5%; Score 93; DB 4; Length 268;
```

Best Local Similarity 94.4%; pred. No. 2.5e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVVCNCFASRNDYS 19
Db 101 LFCNVNDVCNCFASRNDYS 118

RESULT 8

```

US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isoelectric Focusing to
; TITLE OF INVENTION: Modify Cell and Tissue Properties
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/09589927
; CURRENT FILING DATE: 2000-05-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

```

Query Match 83.6%; Score 92; DB 4; Length 260;
Best Local Similarity 88.9%; pred. No. 3.3e-06;
Matches 16; Conservative 1; Mismatches 1; Indels

Qy . 2 LFCNVVCNCFASRNDYS 19
|||:|||||
Db 94 LFCNINNVVCNCFASRNDYS 111

RESULT 9

```

US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolectin B4 Conjugated Cells to Isolate Endothelial Cells
; FILE REFERENCE: 94525-I
; CURRENT APPLICATION NUMBER: US/09/277-665-2
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2

```

Query Match 83.6%; Score 92; DB 4; Length 260;
Best Local Similarity 88.9%; Pred. No. 3.3e-06;
Matches 16; Conservative 1; Mismatches 1; Indels

Qy 2 LFCNVNVCNCFASRNDYS 19
|||:|||||
Db 94 LFCNINNVNVCNCFASRNDYS 111

RESULT 10

US-09-589-987-2
; Sequence 2, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251

; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07

```

; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260

```

US-09-589-987-2

Query Match 83.6%; Score 92; DB 4; Length 260;
Best Local Similarity 88.9%; Pred. No. 3.3e-06;
Matches 16; Conservative 1; Mismatches 1; Indels

Qy 2 LFCNVVCNCFASRNDYS 19
|||:|||||||
Db 94 LFCNINNVVCNCFASRNDYS 111

RESULT 11

```

US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Ty
; FILE OF INVENTION: 951263A
; CURRENT APPLICATION NUMBER: US/08/39
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/62109
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Cal
US-08-399-889-24

```

Query Match 83.6%; Score 92; DB 2; Length 471;
Best Local Similarity 88.9%; Pred. No. 6.1e-06;
Matches 16: Conservative 1; Mismatches 1; Indels

Qy 2 LFCNVVCNCFASRNDYS 19
|||:|||||
pb 304 LFCNINDVCNCFASRNDYS 321

RESULT 12

```

US-09-167-364-24
; Sequence 34, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain T
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/1
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/3998
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

```

Query Match 83.6%; Score 92; DB 3; Length 471;

Best Local Similarity 88.9%; Pred. No. 6.1e-06; Mismatches 1; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LFCNVNVCNCFASRNDYS 19
Db 304 LFCNINDVCNCFASRNDYS 321

RESULT 13

US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 83.6%; Score 92; DB 3; Length 471;
Best Local Similarity 88.9%; Pred. No. 6.1e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LFCNVNVCNCFASRNDYS 19
Db 304 LFCNINDVCNCFASRNDYS 321

RESULT 14

US-09-589-927-10
; Sequence 10, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525I
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-10

Query Match 81.8%; Score 90; DB 4; Length 264;
Best Local Similarity 83.3%; Pred. No. 6.7e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 LFCNVNVCNCFASRNDYS 19
Db 98 MFCNINNVCNCFASRNDYS 115

RESULT 15

US-09-277-665-10
; Sequence 10, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-I
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-10

Query Match 81.8%; Score 90; DB 4; Length 264;
Best Local Similarity 83.3%; Pred. No. 6.7e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 98 MFCNINNVCNCFASRNDYS 115

Search completed: April 5, 2004, 07:07:26
Job time : 5.46247 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.5569 Seconds
(without alignments)
467.378 Million cell updates/sec

Title: US-10-032-221B-42
Perfect score: 145
Sequence: 1 QRFTTMEFLFDVNDVNFASRNDYS 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:
1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 122 | 84.1 | 220 | 2 B49736 | collagen alpha 3(I) |
| 2 | 122 | 84.1 | 1670 | 1 CGHU3B | collagen alpha 3(I) |
| 3 | 121 | 83.4 | 471 | 2 A39024 | collagen alpha 3(I) |
| 4 | 116 | 80.0 | 161 | 2 A49488 | collagen alpha 3(I) |
| 5 | 116 | 80.0 | 246 | 2 I48302 | collagen alpha 3(I) |
| 6 | 106 | 73.1 | 253 | 2 I48304 | collagen alpha 5(I) |
| 7 | 106 | 73.1 | 754 | 2 A53267 | collagen alpha 5(I) |
| 8 | 106 | 73.1 | 1691 | 1 S22917 | collagen alpha 5(I) |
| 9 | 105 | 72.4 | 258 | 2 B61228 | collagen alpha 1(I) |
| 10 | 105 | 72.4 | 1669 | 1 CGHU4B | collagen alpha 1(I) |
| 11 | 105 | 72.4 | 1669 | 1 CGMS4B | collagen alpha 1(I) |
| 12 | 96 | 66.2 | 1747 | 2 A54121 | collagen alpha 1(I) |
| 13 | 96 | 66.2 | 1752 | 2 A45107 | collagen alpha 1(I) |
| 14 | 91 | 62.8 | 1758 | 2 T29350 | hypothetical prote |
| 15 | 91 | 62.8 | 1759 | 2 T29351 | collagen alpha 2(I) |
| 16 | 91 | 62.8 | 1753 | 2 S16366 | collagen alpha 2(I) |
| 17 | 87 | 60.0 | 281 | 2 A34476 | collagen alpha 2(I) |
| 18 | 83 | 57.2 | 1744 | 2 S40991 | collagen alpha 1(I) |
| 19 | 81 | 55.9 | 1691 | 1 CGHU6B | collagen alpha 6(I) |
| 20 | 78 | 53.8 | 775 | 2 A61228 | collagen alpha 2(I) |
| 21 | 78 | 53.8 | 1707 | 2 A33526 | collagen alpha 2(I) |
| 22 | 78 | 53.8 | 1712 | 1 CGHU2B | collagen alpha 2(I) |
| 23 | 65 | 44.8 | 1761 | 2 T13990 | collagen type IV a |
| 24 | 64 | 44.1 | 312 | 2 I48303 | collagen alpha 4(I) |
| 25 | 64 | 44.1 | 623 | 2 A45137 | collagen alpha 4(I) |
| 26 | 64 | 44.1 | 1690 | 1 CGHU1B | collagen alpha 4(I) |
| 27 | 64 | 44.1 | 1775 | 2 A31893 | collagen alpha 1(I) |
| 28 | 63 | 43.4 | 453 | 2 S18804 | collagen alpha 4(I) |
| 29 | 56 | 38.6 | 332 | 2 P82140 | C4-dicarbonylate-b |

30 55 37.9 331 2 H83000 probable C4-dicarb
31 52 35.9 457 2 T23494 phenylalanine 4-mo
32 51.5 35.5 610 2 C70126 DNA mismatch repai
33 50 34.5 331 2 A83534 probable C4-dicarb
34 47.5 32.8 334 2 C71718 hypothetical prote
35 47.5 32.8 490 2 T24497 hypothetical prote
36 47 32.4 252 2 S50806 hypothetical prote
37 46.5 32.1 334 2 B97715 hypothetical prote
38 46 31.7 310 2 T01266 starch synthase DU
39 46 31.7 351 2 JCS904 major capsid prote
40 46 31.7 364 2 A97335 probable membrane
41 46 31.7 795 2 AF2444 hypothetical prote
42 46 31.7 1124 2 F71719 hypothetical prote
43 46 31.7 1462 2 T42639 glucocorticoid rec
44 46 31.7 1463 2 T30193 nuclear receptor C
45 46 31.7 1674 2 T01265 starch synthase DU

ALIGNMENTS

RESULT 1
B49736
collagen alpha 3(IV) chain, medium splice form - human (fragment)
N:Contains: collagen alpha 3(IV) chain, splice form GP-V
C:Species: Homo sapiens (man)
C>Date: 03-May-1994 #sequence revision 12-Nov-1999 #text_change 17-Mar-2000
C:Accession: B49736; D49736; S69111
R:Feng, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A:Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.
A:Reference number: A49736; MUID:94124597; PMID:8294492
A:Accession: B49736
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 169-220 <PEN1>
A:Accession: D49736
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 22-220 <PEN2>
A:Cross-references: GB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107
A:Note: This is the conceptual translation of the nucleic acid submitted to GenBank
R:Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; W.
Eur. J. Biochem. 229, 754-760, 1995
A:Title: Characterization and expression of multiple alternatively spliced transcripts
A:Note: This is the conceptual translation of the nucleic acid submitted to GenBank
A:Reference number: S69111; MUID:95278230; PMID:7758473
A:Accession: S69111
A:Molecule type: mRNA
A:Residues: 1-45,169-204,'L',206-220 <PEN>
C:Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-2q37
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrac
F:1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pre
F:1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status
F:22-220/Domin: carboxyl-terminal nonhelical, NC1 <NC1>
F:324-134/Domin: collagen IV carboxyl-terminal repeat <Ctrl>

Query Match 84.1% Score 122; DB 2; Length 220;
Best Local Similarity 92.3%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTTMEFLFDVNDVNFASRNDYS 27

DB 78 QRFTTMEFLFDVNDVNFASRNDYS 103

RESULT 2
CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human
 N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
 C/Species: Homo sapiens (man)
 C/Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
 C/Accession: A54763; A43928; A44043; A45971; A39786
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Readers, S.T.
 J. Biol. Chem. 269, 23013-23017, 1994
 A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
 A/Reference number: A54763; MUID:94364994; PMID:8083201
 A/Accession: A54763
 A/Molecule type: mRNA
 A/Residues: 1-1670 <MAR>
 A/Cross-references: GB:X80031; NID:G577563; PID:G577564
 A/Experimental source: kidney
 R/Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
 J. Clin. Invest. 89, 592-601, 1992
 A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha
 A/Reference number: A43928; MUID:92147878; PMID:1737849
 A/Accession: A43928
 A/Molecule type: mRNA
 A/Residues: 1331-1524, 'I', 1526-1670 <TUR>
 A/Cross-references: GB:M61379
 A/Experimental source: kidney
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 267, 19780-19784, 1992
 A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
 ction.
 A/Reference number: A44043; MUID:93015826; PMID:1400291
 A/Accession: A44043
 A/Molecule type: DNA, mRNA
 A/Residues: 1386-1670 <QUI>
 A/Cross-references: GB:M2993; NID:G177895; PIDN:AAA1610.1; PID:G177896
 A/Note: sequence extracted from NCBI backbone (NCBIP:1115597)
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 269, 17358, 1994
 A/Reference number: A44738; MUID:94274734; PMID:8005044
 A/Contents: annotation; erratum; correction to intronic sequence in A44043
 R/Bernal, D.; Quinones, S.; Saus, J.
 J. Biol. Chem. 268, 12090-12094, 1993
 A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A/Reference number: A45971; MUID:93280184; PMID:8505332
 A/Accession: A45971
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1427-1444 <BER>
 A/Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly ident
 R/Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Readers, S.T.
 Am. J. Hum. Genet. 49, 545-554, 1991
 A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
 A/Reference number: A39786; MUID:91335370; PMID:1882840
 A/Accession: A39786
 A/Molecule type: mRNA
 A/Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A/Cross-references: GB:S5790; NID:G234418; PIDN:AA19637.1; PID:G234419
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
 C/Genetics:
 A/Gene: GDB:COL4A3
 A/Cross-references: GDB:128351; OMIM:120070
 A/Map position: 2q36-2q37
 A/Introns: 1385/1; 1418/1; 1486/1; 1547/2; 1585/3; 1643/2 #status incomplete
 A/Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
 er associations in the interrupted helical domain (with disulfide and desmosine cross-li
 C/Function:
 A/Description: minor structural component of extracellular basement membrane in kidney
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:28-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
 F:23-42/Domain: amino-terminal nonhelical, NH1 <NHI>

F:43-1438/Region: interrupted helical
 F:791-793/Region: cell attachment (R-G-D) motif
 F:996-998/Region: cell attachment (R-G-D) motif
 F:1154-1156/Region: cell attachment (R-G-D) motif
 F:1306-1308/Region: cell attachment (R-G-D) motif
 F:1345-1347/Region: cell attachment (R-G-D) motif
 F:1432-1434/Region: cell attachment (R-G-D) motif
 F:1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NCL1>
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:31.33.39.41.125.422.476.479.682.722.809.1387/Disulfide bonds: interchain #status pred
 F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1458-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 84.1%; Score 122; DB 1; Length 1670;
 Best Local Similarity 92.3%; Pred. No. 7.7e-10;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVNFASRNDYS 27
 DB 1495 QRFTHMPFLFDNVNDVNFASRNDYS 1520

RESULT 3
 A39024
 C:collagen alpha 3(IV) chain - bovine (fragment)
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
 C/Accession: A39024; S26672; S17802; A35167; C39419; S13747; S20815
 R/Morrison, K.E.; Germino, G.G.; Readers, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A/Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
 A/Reference number: A39024; MUID:91093146; PMID:1985905
 A/Accession: A39024
 A/Molecule type: mRNA
 A/Residues: 1-471 <MOR>
 A/Cross-references: EMBL:MG3139; NID:G162886; PIDN:AAA62708.1; PID:G162887
 R/Butkowski, R.J.; Wieslander, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A/Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A/Reference number: S18432; MUID:87222419; PMID:2438283
 A/Accession: S20672
 A/Molecule type: protein
 A/Residues: 227-228, 'X', 230-244 <BUT>
 R/Saus, J.; Wieslander, J.P.M.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A/Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
 A/Reference number: S17802; MUID:88330844; PMID:3417661
 A/Accession: S17802
 A/Molecule type: protein
 A/Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R/Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 265, 5466-5469, 1990
 A/Title: Glomerular basement membranes. Identification of a fourth chain, alpha4, of type
 A/Reference number: A35167; MUID:90202779; PMID:2318822
 A/Accession: A35167
 A/Molecule type: protein
 A/Residues: 236-258 <GUN>
 R/Gunwar, S.; Ballister, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; NC
 J. Biol. Chem. 266, 15318-15324, 1991
 A/Title: Glomerular basement membranes. Identification of dimeric subunits of the noncol
 A/Reference number: A39419; MUID:91332055; PMID:1869555
 A/Accession: C39419
 A/Molecule type: protein
 A/Residues: 236-255 <GU2>
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;
 F:1-238/Domain: collagenous (fragment) #status predicted <COL>
 F:239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NCL1>
 F:239-353/Domain: repeat NC1 #status predicted <NCL1>
 F:354-471/Domain: repeat NC1 #status predicted <NCL2>

F:232,238/Modified site: hydroxyproline (Pro) #status experimental
F:306-312,417-423/Disulfide bonds: #status predicted

Query Match 83.4%; Score 121; DB 2; Length 471;
Best Local Similarity 88.5%; Pred. No. 2.5e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
|||||:|||||:|||||:|||||:|||||:
DB 296 QRETTMPFLFCNNDVNCVFASRNDYS 321
|||||:|||||:|||||:|||||:|||||:

RESULT 4

S49488
collagen alpha 3(IV) chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C:Accession: S49488
R:Oberbaumer, I.
submitted to the EMBL Data Library, October 1994
A:Description: Cloning of the NC1 domains fo the minor collagen IV chains of mouse via F
ells.

A:Reference number: S49487
A:Accession: S49488
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-161 <OBE>
A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G559916
C:Superfamily: collagen alpha 1(IV) chain

Query Match 80.0%; Score 116; DB 2; Length 161;
Best Local Similarity 84.6%; Pred. No. 4.1e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
|||||:|||||:|||||:|||||:|||||:
DB 4 QRETTMPFLFCNNDVNCVFASRNDYS 29
|||||:|||||:|||||:|||||:|||||:

RESULT 5

I48302
collagen alpha 3(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
C:Accession: I48302; S47278
R:Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
A:Reference number: A54979; MUID:95050957; PMID:7962065
A:Accession: I48302
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <RES>
A:Cross-references: EMBL:X35166; NID:G535197; PID:G535198
C:Superfamily: collagen alpha 1(IV) chain

Query Match 80.0%; Score 116; DB 2; Length 246;
Best Local Similarity 84.8%; Pred. No. 6.8e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
|||||:|||||:|||||:|||||:|||||:
DB 71 QRETTMPFLFCNNDVNCVFASRNDYS 96
|||||:|||||:|||||:|||||:|||||:

RESULT 6

I48304
collagen alpha 5(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999
C:Accession: I48304; S47280
R:Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq
A:Reference number: A54979; MUID:95050957; PMID:7962065
A:Accession: I48304
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <RES>
A:Cross-references: EMBL:X35168; NID:G535201; PIDN:CAA84531.1; PID:G535202
C:Superfamily: collagen alpha 1(IV) chain

Query Match 73.1%; Score 106; DB 2; Length 253;
Best Local Similarity 73.1%; Pred. No. 2.2e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
|||||:|||||:|||||:|||||:|||||:
DB 79 RRFSTMPFMCNINNVCNPFASRNDYS 104
|||||:|||||:|||||:|||||:|||||:

RESULT 7

A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267
R:Zheng, K.; Thorne, P.S.; Marrano, P.; Baumal, R.; McInnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-
en type IV.
A:Reference number: A55267; MUID:94224868; PMID:8171024
A:Accession: A55267
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-754 <ZHE>
A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548
C:Superfamily: collagen alpha 1(IV) chain

Query Match 73.1%; Score 106; DB 2; Length 754;
Best Local Similarity 73.1%; Pred. No. 7.9e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
|||||:|||||:|||||:|||||:|||||:
DB 587 RRFSTMPFMCNINNVCNPFASRNDYS 612
|||||:|||||:|||||:|||||:|||||:

RESULT 8

S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N:Alternate names: procollagen alpha 5(IV) chain
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A
R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and iden
n Alport syndrome patient.

A:Reference number: S22917; MUID:92316923; PMID:1352287
A:Accession: S22917
A:Molecule type: mRNA
A:Residues: 1-967 <ZHO>
A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234
R:Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A:Title: Structure of the human type IV collagen COL4A5 gene.
A:Reference number: A54365; MUID:94165049; PMID:8120014
A:Accession: A54365
A:Molecule type: DNA
A:Residues: 1-922 <ZHO>
A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAC27816.1; P
R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggv
Science 261, 1167-1169, 1993
A:Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited

A;Cross-references: GB:S75903; NID:g913882; PIDN:AAB33374.1; PID:g913883
A;Note: perimature termination mutation from a patient with Alport syndrome; one other
R;Lemlink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.
Genomics 17, 485-489, 1993
A;Title: Identification of four novel mutations in the COL4A5 gene of patients with Alj
A;Reference number: I54188; MUID:94010948; PMID:8406498
A;Accession: I54188
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1604-1607,'VHDAAYKC' <LEM>
A;Cross-references: GB:S65767; NID:g425563; PIDN:AAAD13967.1; PID:g4261667
A;Note: frameshift mutation from a patient with Alport syndrome; five other mutations
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A5; AYS
A;Cross-references: GDB:120596; OMIM:303630
A;Map position: Xq22-Xq22
A;Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 215/3; 221/3/
3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1;
A;Notes: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands wi
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 1
ong trimer amino-terminal domains [with disulfide and desmosine cross-links], dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C;Function:
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; gly
F;1-26/Domain: signal sequence #status predicted <Sig>
F;27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MA
F;27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #statu
F;27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F;27-1462/Region: interrupted helical
F;1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1473-1691/Domain: collagen IV carboxyl-terminal repeat <CrI>
F;1583-1687/Domain: collagen IV carboxyl-terminal repeat <CrI>
F;29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F;125/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicte
F;1527-1533,1638-1644/Disulfide bonds: #status predicted
F;1592-1684,1628-1687/Disulfide bonds: (or 1592-1687, 1628-1684) #status predicted

Query Match 73.1%; Score 106; DB 1; Length 1691;
Best Local Similarity 73.1%; Pred. No. 2e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTWPELFNNVDVNFASRNDS 27
:::|||||:|:|:|||||||

Ds 1517 RPFSTMPFFCNINNVCFASRNDS 1542
:::|||||:|:|:|||||||

RESULT 9
B61228
collagen alpha 1(IV) chain - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
C;Accession: B61228
R;Yamauchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991
A;Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium
A;Reference number: A61228; MUID:92010685; PMID:1717398
A;Accession: B61228
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-258 <AM>
C;Superfamily: collagen alpha 1(IV) chain

Query Match 72.4%; Score 105; DB 2; Length 258;
Best Local Similarity 73.1%; Pred. No. 3.2e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTWPELFNNVDVNFASRNDS 27
:::|||||:|:~|:|||||||

Db 84 RKFTWPELFNCNNVCFASRNDYS 109

RESULT 10

CGHU4B
collagen alpha 1(IV) chain precursor - human
N;Alternate names: procollagen alpha 1(IV) chain
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
A;Reference number: S16876; MUID:89340433; PMID:2701944
A;Accession: S16876
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1669 <SO1>
A;Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803
R;Soininen, R.; Huotari, M.; Hosikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A;Reference number: A92650; MUID:89034231; PMID:3182844
A;Accession: A32117
A;Molecule type: DNA
A;Residues: 1-28 <SO12>
A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233
R;Roeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c
A;Reference number: S02738; MUID:89030632; PMID:2846280
A;Accession: S02738
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6, 'L', 8-28 <POE>
A;Cross-references: EMBL:X12784; NID:G30072
R;Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
A;Reference number: S00048; MUID:88029471; PMID:3311751
A;Accession: S00048
A;Molecule type: mRNA
A;Residues: 1-318, 'A', 320-944 <BRAL>
A;Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067
A;Accession: S25826
A;Molecule type: protein
R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S
A;Reference number: A23115; MUID:86004708; PMID:4043082
A;Accession: A23115
A;Molecule type: protein
A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>
A;Experimental source: Placenta
A;Note: the amino end of the mature form is blocked
R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
FEBS Lett. 225, 188-194, 1987
A;Title: Complete primary structure of the alpha1(1)-chain of human basement membrane (ty
A;Reference number: S00207; MUID:88083584; PMID:3691802
A;Accession: S00207
A;Molecule type: mRNA
A;Residues: 244-530 <SO13>
A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549
R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
A;Reference number: S39614; MUID:94038963; PMID:8223488
A;Accession: S39614
A;Molecule type: protein
A;Residues: 371-554 <EBL>
R;Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984

A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid
A;Reference number: A02863; MUID:85003629; PMID:6434307
A;Accession: A02863
A;Molecule type: protein
A;Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 99
A;Experimental source: Placenta
R;Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A;Title: Peptin fragments of human placental basement-membrane collagens showing inter
A;Reference number: S16908; MUID:82005835; PMID:6792033
A;Accession: A58517
A;Molecule type: protein
A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'XX', 1408-1409, 'X', 1411-
R;MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Pieszek, P.P.
Biochemistry 22, 4940-4948, 1983
A;Title: Isolation and characterization of pepsin-solubilized human basement membrane
A;Reference number: S16910; MUID:84053346; PMID:6416291
A;Accession: S16910
A;Molecule type: protein
A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'V', 946, 'X', 94
A;Experimental source: Placenta
R;Pillayantem, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;
J. Biol. Chem. 260, 7681-7687, 1985
A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
A;Reference number: S01466; MUID:85207819; PMID:2581969
A;Accession: S01466
A;Molecule type: mRNA
A;Residues: 1256-1669 <PIH>
A;Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424
R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A;Title: Restricted homology between human alpha-1 type IV and other procollagen chain
A;Reference number: S16879; MUID:85216555; PMID:2582422
A;Accession: S16879
A;Molecule type: mRNA
A;Residues: 1259-1669 <BRI>
A;Cross-references: EMBL:M1315; NID:G180817; PIDN:AAA52042.1; PID:G180818
R;Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss
Eur. J. Biochem. 147, 217-224, 1985
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha
A;Reference number: A02864; MUID:85127033; PMID:2578961
A;Accession: S19091
A;Molecule type: protein
A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491; 1501-1514, 'X', 1516-1519; 1534-1553, 'X'
R;Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyte:
A;Reference number: S02550; MUID:8905112; PMID:2844531
A;Contents: annotation; disulfide bonds
C;Genetics:
A;Gene: GDB:COL4A1
A;Cross-references: GDB:119791; OMIM:120130
A;Map position: 13q34-13q34
A;Intons: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 23;
A;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 1;
C;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha
C;Trimer associations in the interrupted helical domain (with disulfide and desmosine)
C;Description: structural component of extracellular basement membrane
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicati
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F;29-162/Domain: amino-terminal nonhelical, 7S <7SD>
F;163-1440/Domain: interrupted helical <COL>
F;414-452/Region: integrin binding #status experimental
F;597-599/Region: cell attachment (R-G-D) motif
F;917-919/Region: cell attachment (R-G-D) motif
F;968-970/Region: cell attachment (R-G-D) motif
F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <Ctri>

R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
Gene 43, 301-304, 1986
A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligode-
A;Reference number: A25636; MUID:86301886; PMID:3755692
A;Accession: A25636
A;Molecule type: mRNA
A;Residues: 1149-1396, 'S', 1398-1424 <NAT>
A;Cross-references: EMBL:M14042; NID:G19286; PIDN:AAA37342.1; PID:G19287
A;Note: the authors translated the codon CAG for residue 1374 as Arg
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihl:
J. Biol. Chem. 262, 8496-8499, 1987
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
A;Reference number: A34680; MUID:87250460; PMID:3597383
A;Accession: A23301
A;Molecule type: mRNA
A;Residues: 1441-1669 <KUR>
A;Cross-references: EMBL:M15832; NID:G19282; PIDN:AAA37340.1; PID:G387115
R;Killen, P.D.; Burbelo, P.D.; Martin, G.R.; Yamada, Y.
J. Biol. Chem. 263, 12310-12314, 1988
A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequen-
A;Reference number: S19079; MUID:88315019; PMID:2842328
A;Accession: S19079
A;Molecule type: DNA
A;Residues: 1-28 <K12>
A;Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
J. Biol. Chem. 263, 19274-19277, 1988
A;Title: Head-to-head arrangement of murine type IV collagen genes.
A;Reference number: A32702; MUID:89066738; PMID:3198626
A;Accession: A32003
A;Molecule type: DNA
A;Residues: 1-28 <KAY>
A;Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449
R;Burbelo, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A., 85, 3679-3682, 1988
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional pro-
A;Reference number: A34220; MUID:89071759; PMID:3200851
A;Accession: A31766
A;Molecule type: DNA
A;Residues: 1-28 <BUR>
A;Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668
R;Sakurai, Y.; Sullivan, M.; Yamada, Y.
J. Biol. Chem. 261, 6654-6657, 1986
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen gene
A;Reference number: S19094; MUID:86196099; PMID:3009468
A;Accession: S19094
A;Molecule type: DNA
A;Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>
R;Schuppan, D.; Timpl, R.; Glanville, R.W.
FEBS Lett. 115, 297-300, 1980
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane
A;Reference number: S16903; MUID:80246483; PMID:6772473
A;Accession: S16903
A;Molecule type: protein
A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-1
R;Schuppan, D.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 123, 505-512, 1982
A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial am
A;Reference number: A25991; MUID:82186723; PMID:6804236
A;Accession: A25991
A;Molecule type: protein
A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 1
61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-111
A;Accession: B25991
A;Molecule type: protein
A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1233
3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282; 1316-1318, 'X', 1320-1327, 'X', 13
R;Weber, S.; Engel, J.; Whedemann, H.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A;Title: Subunit structure and assembly of the globular domain of basement-membrane col
A;Reference number: S17801; MUID:84132058; PMID:6698021
A;Accession: S17801

RESULT 13
A45407
collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 22-Sep-1993 #sequence revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A45407; A43903; A33940
E:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.
J. Biol. Chem. 268, 5249-5254, 1993
A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and analysis of the alpha 1(IV) chain
A:Reference number: A45407; PMID:93186842; PMID:8444899

```

Qy      2 QRETTMPFLFDNVNDVDFASRNDYS 27
      |||:|||||:|||||
Db      1581 QRESIMPFLFCDFNVCNYASRNDKS 1606

```

T29351
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans
N;Alternate names: collagen alpha 2(IV) chain precursor clb-1
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Feb-2000
C;Accession: T29351
R;Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid F01G12.
A;Reference number: Z20611
A;Accession: T29351
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1759 <WUX>
A;Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A;Experimental source: strain Bristol N2; clone F01G12
C;Genetics:
A;Gene: CESP:F01G12.5a
A;Map position: X
A;Intons: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3
C;Superfamily: collagen alpha 1(IV) chain

Query Match 62.8%; Score 91; DB 2; Length 1759;
Best Local Similarity 69.2%; Pred. NO. 3.8e-05;
Matches 18; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
|||:|||||:|:|||||
Db 1582 QRETTMPFLFDNVNDVNFASRNDKS 1607

Search completed: April 5, 2004, 07:05:39
Job time : 5.5569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.3952 Seconds
(without alignments)
413.557 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTMRPLFDVNDVNFASGRDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 122 | 84.1 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 2 | 121 | 83.4 | 471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 3 | 106 | 73.1 | 754 | 1 CA54 CANFA | P29400 canis faml |
| 4 | 106 | 73.1 | 1685 | 1 CA54 HUMAN | P29400 homo sapien |
| 5 | 105 | 72.4 | 1669 | 1 CA14 HUMAN | P02462 mus musculus |
| 6 | 105 | 72.4 | 1669 | 1 CA14 MOUSE | P27393 ascaris suu |
| 7 | 91 | 62.8 | 1763 | 1 CA24 ASCSU | P17140 caenorhabdi |
| 8 | 87 | 60.0 | 1758 | 1 CA24 CAEEL | P17139 caenorhabdi |
| 9 | 83 | 57.2 | 1758 | 1 CA14 CAEEL | Q14031 homo sapien |
| 10 | 81 | 55.9 | 1691 | 1 CA64 HUMAN | P08122 mus musculus |
| 11 | 78 | 53.8 | 1707 | 1 CA24 MOUSE | P08572 homo sapien |
| 12 | 78 | 53.8 | 1712 | 1 CA24 HUMAN | P55787 cryctolagus |
| 13 | 64 | 44.1 | 623 | 1 CA44 RABIT | P53420 homo sapien |
| 14 | 64 | 44.1 | 1690 | 1 CA44 HUMAN | P08120 drosophila |
| 15 | 64 | 44.1 | 1775 | 1 CA14 DROME | Q29442 bos taurus |
| 16 | 63 | 43.4 | 453 | 1 CA44 BOVIN | O48093 liasis macu |
| 17 | 52 | 35.9 | 371 | 1 CYB LIAPA | O48098 liasis papu |
| 18 | 52 | 35.9 | 371 | 1 CYB LIAPA | P09025 caenorhabdi |
| 19 | 52 | 35.9 | 457 | 1 PH4H CAEEL | P09129 bifidobacte |
| 20 | 51.5 | 35.5 | 610 | 1 MUTL BORBU | Q48100 loxocemus b |
| 21 | 49 | 33.8 | 371 | 1 CYB LOXBI | Q95400 dipodomys n |
| 22 | 48 | 33.1 | 379 | 1 CYB DIPNE | Q95400 dipodomys n |
| 23 | 47.5 | 32.6 | 334 | 1 Y092 RICPR | Q95400 dipodomys n |
| 24 | 47 | 32.4 | 252 | 1 YUG6 YEAST | P40364 saccharomyc |
| 25 | 47 | 32.4 | 371 | 1 CYB ERYJA | O48076 eryx jaculu |
| 26 | 47 | 32.4 | 825 | 1 XFP BIFAN | Q9AEM9 bifidobacte |
| 27 | 46 | 31.7 | 218 | 1 IF6 METAC | Q8TIN4 methanosarc |
| 28 | 46 | 31.7 | 328 | 1 NODZ AZOCA | Q43966 azorhizobiu |
| 29 | 46 | 31.7 | 371 | 1 CYB COLCO | Q9M110 coluber con |
| 30 | 46 | 31.7 | 371 | 1 CYB ELANI | Q9M110 coluber con |
| 31 | 46 | 31.7 | 1462 | 1 NCO2 MOUSE | Q61026 mus musculu |
| 32 | 46 | 31.7 | 1464 | 1 NCO2 HUMAN | Q15596 homo sapien |
| 33 | 46 | 31.7 | 1465 | 1 NCO2 RAT | Q9WUI9 rattus norv |

RESULT 1

CA34 HUMAN
ID CA34 HUMAN STANDARD; PRT; 1670 AA.
AC Q01955; Q9BOT2; 125 1 YOM2 PHOPR
DT 01-OCT-1996 (Rel. 34, Created) 288 1 T2D2_STRPN
DT 15-JUL-1999 (Rel. 38, Last sequence update) 1397 1 CID_DROME
DT 10-OCT-2003 (Rel. 42, Last annotation update) 370 1 CYB_MICIK
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen). 430 1 RUMA_SALTI
GN COL4A3. 430 1 RUMA_SALTY
OS Homo sapiens (Human). 432 1 RUMA_ECOL6
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; 432 1 RUMA_ECOLI
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 432 1 RUMA_ECOLI
OX NCBI_Taxid=9606; 432 1 RUMA_ECOLI
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. Biol. Chem. 269:23013-23017(1994).
RN [2]

REVIEWS

RA Leinonen A.;
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
RP CYS-1661; AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
RP PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064696; PubMed=11134255;
RA Heidet L., Arondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
RN [4]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:19780-19784(1992).
RN [5]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1882840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;

ALIGNMENTS

34 45.5 31.4 125 1 YOM2 PHOPR
35 45.5 31.4 288 1 T2D2_STRPN
36 45.5 31.4 1397 1 CID_DROME
37 45 31.0 370 1 CYB_MICIK
38 45 31.0 430 1 RUMA_SALTI
39 45 31.0 430 1 RUMA_SALTY
40 45 31.0 432 1 RUMA_ECOL6
41 45 31.0 432 1 RUMA_ECOLI
42 45 31.0 432 1 RUMA_ECOLI
43 45 31.0 688 1 SP12 YEAST
44 45 31.0 2291 1 SPCB_DROME
45 45 31.0 2301 1 POLG_TWEVD

P29740 photobacter
P09357 streptococc
P19538 drosophila
Q9MIK2 micropechis
Q8Z446 salmonella
Q8ZNE1 salmonella
Q8XED8 escherichia
Q8FEG6 escherichia
P55335 escherichia
P13574 saccharomyc
Q00963 drosophila
P13899 t genome po

Query Match 84.1%; Score 122; DB 1; Length 1670;
 Best Local Similarity 92.3%; Pred. No. 6.4e-10;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27
 |||||
 Db 1495 QRTTTPFLFDVNDVNFASRNDYS 1520

RESULT 2

CA34_BOVIN STANDARD; PRT; 471 AA.
 AC Q28084;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 3(IV) chain (Fragment).
 GN COL4A3.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RX MEDLINE=91093146; PubMed=1985905;
 RA Morrison K.E., Germino G.G., Readers S.T.;
 RT "Use of the polymerase chain reaction to clone and sequence a cDNA
 encoding the bovine alpha 3 chain of type IV collagen.";
 RL J. Biol. Chem. 266:34-39(1991).
 RN [2]
 RP SEQUENCE OF 227-258.
 RC TISSUE=Kidney;
 RX MEDLINE=90202779; PubMed=2318822;
 RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
 RT "Glomerular basement membrane. Identification of a fourth chain,
 alpha 4, of type IV collagen.";
 RL J. Biol. Chem. 265:5466-5469(1990).
 RN [3]
 RP SEQUENCE OF 227-254.
 RX MEDLINE=88330844; PubMed=3417661;
 RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
 RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
 of collagen IV.";
 RL J. Biol. Chem. 263:13374-13380(1988).
 RN [4]
 RP SEQUENCE OF 227-244.
 RX MEDLINE=87222419; PubMed=2438283;
 RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
 RA Hudson B.G.;
 RT "Localization of the Goodpasture epitope to a novel chain of basement
 membrane collagen.";
 RL J. Biol. Chem. 262:7874-7877(1987).
 CC -1- FUNCTION: Type IV collagen is the major structural component of
 glomerular basement membranes (GBM), forming a 'chicken-wire'
 meshwork together with laminins, proteoglycans and entactin/
 nidogen.
 CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 alpha 6(IV), each of which can form a triple helix structure
 with 2 other chains to generate type IV collagen network.
 CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the
 G-X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type

IV collagens.
 -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 EMBL; M63139; AAA62708.1; -;
 PIR; A39024; A39024.
 InterPro; IPR008160; Collagen.
 InterPro; IPR001442; Procollagen_C.
 Pfam; PF01413; C4; 2.
 Pfam; PF01391; Collagen; 4.
 ProDom; PD003923; ProcollagenC4; 1.
 SMART; SM00111; C4; 2.
 Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 NON TPR 1 1
 DOMAIN <1 238 TRIPLE-HELICAL REGION.
 DOMAIN 239 471 NONHELICAL REGION (NC1).
 SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
 MOD RES 232 232 HYDROXYLATION.
 MOD RES 238 238 HYDROXYLATION.
 DISULFID 261 352 OR 349 (BY SIMILARITY).
 DISULFID 294 349 OR 352 (BY SIMILARITY).
 DISULFID 306 312 BY SIMILARITY.
 DISULFID 371 466 OR 463 (BY SIMILARITY).
 DISULFID 405 463 OR 466 (BY SIMILARITY).
 DISULFID 417 423 BY SIMILARITY.
 CONFLICT 253 253 S -> Y (IN REF. 3).
 SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match 83.4%; Score 121; DB 1; Length 471;
 Best Local Similarity 88.5%; Pred. No. 2.1e-10;
 Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27
 |||||
 Db 296 QRTTTPFLFDVNDVNFASRNDYS 321

RESULT 3

CA54_CANFA STANDARD; PRT; 754 AA.
 ID CA54 CANFA
 AC Q28247;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 5(IV) chain (Fragment).
 GN COL4A5.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Samoyed; TISSUE=Kidney;
 RX MEDLINE=94224868; PubMed=8171024;
 RA Zheng K., Thorne P.S., Marrano P., Bauman R., McInnes R.R.;
 RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
 human X-linked hereditary nephritis resulting from a single base
 mutation in the gene encoding the alpha 5 chain of collagen type
 IV.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
 CC -1- FUNCTION: Type IV collagen is the major structural component of
 glomerular basement membranes (GBM), forming a 'chicken-wire'
 meshwork together with laminins, proteoglycans and entactin/
 nidogen.
 CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-

CC alpha 5(IV), each of which can form a triple helix structure with
 CC 2 other chains to generate type IV collagen network.
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-
 CC x-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -!- DISPAR: A defect in COL4A5 has been found to be the cause of
 CC canine X-linked hereditary nephritis (HN), a disease similar to
 CC that in humans (also referred to as Alport syndrome) characterized
 CC by progressive renal failure and neurosensory deafness.
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U07888; AAB60258.1; -;
 CC FTR: A55267; A55267.
 CC InterPro: IPR008161; Clg_helix.
 CC InterPro: IPR008160; Collagen.
 CC InterPro: IPR001442; Procollagn4_C.
 CC Pfam: PF01413; C4; 2.
 CC Pfam: PF01391; Collagen; 8.
 CC ProDom: PD000007; Clg_helix; 1.
 CC ProDom: PD003923; ProcollagnC4; 1.
 CC SMART: SM00111; C4; 2.
 CC ExPASy: P000007; Clg_helix; 1.
 CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 CC FT NON_TER 1 1
 CC FT DOMAIN <1 530 TRIPLE-HELICAL REGION
 CC FT DISULFID 531 754 NON-HELICAL REGION (NC1).
 CC FT DISULFID 552 643 OR 640 (BY SIMILARITY).
 CC FT DISULFID 585 640 OR 643 (BY SIMILARITY).
 CC FT DISULFID 597 603 BY SIMILARITY.
 CC FT DISULFID 662 ? OR 754 (BY SIMILARITY).
 CC FT DISULFID 696 754 BY SIMILARITY.
 CC FT DISULFID 708 714 BY SIMILARITY.
 CC FT NON_TER 754 754
 CC SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;
 CC
 CC Query Match 73.1%; Score 106; DB 1; Length 754;
 CC Best Local Similarity 73.1%; Pred. No. 6.1e-08;
 CC Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 2 QRTTTFPLDNDVDFNFSRNDYS 27
 CC DB 587 RRFSTPFMFCNNVNFNFSRNDYS 612
 CC
 CC RESULT 4
 CC CAS4_HUMAN STANDARD; PRT; 1685 AA.
 CC ID CAS4_HUMAN
 CC AC P29406; Q16006; Q16126;
 CC DT 01-DEC-1992 (Rel. 24, Created)
 CC DT 01-FEB-1994 (Rel. 28, Last sequence update)
 CC DE 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DT Collagen alpha 5(IV) chain precursor.
 CC GN COL4A5.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94165049; PubMed=8120014;
 RA Zhou J., Leinonen A., Tryggvason K.;
 RT "Structure of the human type IV collagen COL4A5 gene.";
 RL J. Biol. Chem. 269:6608-6614(1994).
 RN [2]
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
 RC TISSUE=Kidney;
 RX MEDLINE=92316923; PubMed=1352287;
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen
 RT chain and identification of a single-base mutation in exon 23
 RT converting glycine 521 in the collagenous domain to cysteine in an
 RT Alport syndrome patient";
 RL J. Biol. Chem. 267:12475-12481(1992).
 RN [3]
 RP SEQUENCE OF 85-1685 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=90337990; PubMed=2380186;
 RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
 RT "Complete primary structure of the triple-helical region and the
 RT carboxyl-terminal domain of a new type IV collagen chain, alpha
 RT 5(IV).";
 RL J. Biol. Chem. 265:13758-13766(1990).
 RN [4]
 RP SEQUENCE OF 924-1685 FROM N.A.
 RX MEDLINE=91169491; PubMed=2004755;
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
 RT "Characterization of the 3' half of the human type IV collagen alpha
 RT 5 gene that is affected in the Alport syndrome.";
 RL Genomics 9:1-9(1991).
 RN [5]
 RP SEQUENCE OF 914-1685 FROM N.A.
 RX MEDLINE=90160375; PubMed=1689491;
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
 RA Tryggvason K.;
 RT "Identification of a distinct type IV collagen alpha chain with
 RT restricted kidney distribution and assignment of its gene to the
 RT locus of X chromosome-linked Alport syndrome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
 RN [6]
 RP SEQUENCE OF 1442-1471 FROM N.A.
 RX MEDLINE=90252791; PubMed=2339699;
 RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,
 RA Sheer D., Solomon E., Pihlajaniemi T.;
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
 RT to the region of the X chromosome containing the Alport syndrome
 RT locus.";
 RL Am. J. Hum. Genet. 46:1024-1033(1990).
 RN [7]
 RP SEQUENCE OF 1-20 FROM N.A.
 RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
 RA Marynen P.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
 RX MEDLINE=94133540; PubMed=8301933;
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
 RA Cassiman J.-J., Marynen P.;
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood
 RT cells: a complex mutation in the COL4A5 gene of an Alport patient
 RT deletes the NC1 domain.";
 RL Kidney Int. 44:1316-1321(1993).
 RN [9]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 RN [10]
 RP VARIANT AS SER-1564.

RL J. Biol. Chem. 264:13565-13571(1989).
RN [2]
RP SEQUENCE OF 46-1257 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88083584; PubMed=3691802;
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
RT "Complete primary structure of the alpha 1-chain of human basement
membrane (type IV) collagen.";
RL FEBS Lett. 225:188-194(1987).
RN [3]
RP SEQUENCE OF 1-943 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88029471; PubMed=3111751;
RA Brazel D., Oberbauer I., Dieringer H., Babel W., Gnanville R.W.,
RT Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
basement membrane collagen (type IV) reveals 21 non-triplet
interruptions located within the collagenous domain.";
RL Eur. J. Biochem. 168:529-536(1987).
RN [4]
RP SEQUENCE OF 28-243;
RX MEDLINE=86004708; PubMed=4043082;
RA Gnanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
RT "Amino acid sequence of the N-terminal aggregation and cross-linking
region (7S domain) of the alpha 1 (IV) chain of human basement
membrane collagen.";
RL Eur. J. Biochem. 152:213-219(1985).
RN [5]
RP SEQUENCE OF 534-1447.
RX MEDLINE=85003629; PubMed=6434307;
RA Babel W., Gnanville R.W.;
RT "Structure of human-basement-membrane (type IV) collagen. Complete
amino acid sequence of a 914-residue-long pepsin fragment from the
alpha 1(IV) chain.";
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85207819; PubMed=2581969;
RA Fihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RT Chong M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
procollagen reveal an unusual homology of amino acid sequences in two
halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
collagen are divergently encoded on opposite DNA strands and have an
overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
carboxyterminal, non-collagenous aggregation and cross-linking domain
of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.

CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -1- PTM: Lysines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in all cases and bind carboxylates.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M26576; AAA53098.1; JOINED.
CC EMBL; J04217; AAA53098.1; JOINED.
CC EMBL; M26550; AAA53098.1; JOINED.
CC EMBL; M26540; AAA53098.1; JOINED.
CC EMBL; M26542; AAA53098.1; JOINED.
CC EMBL; M26543; AAA53098.1; JOINED.
CC EMBL; M26544; AAA53098.1; JOINED.
CC EMBL; M26545; AAA53098.1; JOINED.
CC EMBL; M26546; AAA53098.1; JOINED.
CC EMBL; M26547; AAA53098.1; JOINED.
CC EMBL; M26537; AAA53098.1; JOINED.
CC EMBL; M26538; AAA53098.1; JOINED.
CC EMBL; M26548; AAA53098.1; JOINED.
CC EMBL; M26549; AAA53098.1; JOINED.
CC EMBL; M26551; AAA53098.1; JOINED.
CC EMBL; M26552; AAA53098.1; JOINED.
CC EMBL; M26553; AAA53098.1; JOINED.
CC EMBL; M26554; AAA53098.1; JOINED.
CC EMBL; M26555; AAA53098.1; JOINED.
CC EMBL; M26556; AAA53098.1; JOINED.
CC EMBL; M26557; AAA53098.1; JOINED.
CC EMBL; M26558; AAA53098.1; JOINED.
CC EMBL; M26559; AAA53098.1; JOINED.
CC EMBL; M26560; AAA53098.1; JOINED.
CC EMBL; M26561; AAA53098.1; JOINED.
CC EMBL; M26562; AAA53098.1; JOINED.
CC EMBL; M26563; AAA53098.1; JOINED.
CC EMBL; M26564; AAA53098.1; JOINED.
CC EMBL; M26565; AAA53098.1; JOINED.
CC EMBL; M26566; AAA53098.1; JOINED.
CC EMBL; M26567; AAA53098.1; JOINED.
CC EMBL; M26568; AAA53098.1; JOINED.
CC EMBL; M26569; AAA53098.1; JOINED.
CC EMBL; M26570; AAA53098.1; JOINED.
CC EMBL; M26571; AAA53098.1; JOINED.
CC EMBL; M26572; AAA53098.1; JOINED.
CC EMBL; M26573; AAA53098.1; JOINED.
CC EMBL; M26574; AAA53098.1; JOINED.
CC EMBL; M26575; AAA53098.1; JOINED.
CC EMBL; Y00706; CAA69698.1; JOINED.
CC EMBL; X05561; CAA29075.1; JOINED.
CC EMBL; M10940; AAA52006.1; JOINED.
CC EMBL; M1315; AAA52042.1; JOINED.
CC PIR; S16876; CGHU4B.
CC Genew; HGNC:2202; COL4A1.

```

DR MIM: 120130;
DR InterPro: IPR008161; C1g helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR001442; Procollagen4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 24.
DR ProDom: PD000007; C1g helix; 6.
DR ProDom: PD003923; ProcollagenC4; 1.
DR SMART: SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1662.
FT DISULFID 1570 1665 OR 1662.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1662 OR 1665.
FT CONFLICT 237 238 SG -> KE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
FT SEQUENCE 1669 AA; 160611 MW; 3EBBA6DFB9B8A84 CRC64;

Query Match 72.4%; Score 105; DB 1; Length 1669;
Best Local Similarity 73.1%; Pred. No. 2.1e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTHPFLFDNNVDNPFASNDYS 27
DB 1495 RKFTMPFLFCNNVNCNPFASNDYS 1520

RESULT 6
CAI4_MOUSE
ID CAI4_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen

```

```

RT from mouse.";
RL REBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PMW: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PMW: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; J03758; AAA37439.1; -
 DR EMBL; M23333; AAA51625.1; -
 DR EMBL; J04594; AAA50292.1; -
 DR EMBL; X06777; CAA29946.1; -
 DR EMBL; X02201; CAA26132.1; -
 DR EMBL; M15832; AAA37340.1; -
 DR EMBL; M14042; AAA37341.1; -
 DR EMBL; M12879; AAA37343.1; -
 DR EMBL; M13024; -; NOT_ANNOTATED_CDS.
 DR EMBL; M13025; -; NOT_ANNOTATED_CDS.
 DR EMBL; M13026; AAA37344.1; -
 DR EMBL; M13027; AAA37345.1; -
 DR EMBL; M13043; AAA37346.1; -
 DR EMBL; J04448; AAA37437.1; -
 DR PIR; A33525; CGMS4B.
 DR MGD; MGI-88454; Col4a1.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD000007; C1g_helix; 6.
 DR ProDom; PD003923; Procollagnc4; 1.
 DR SMART; SM00111; C4; 2.
 DR Repeat; Hydroxylation; Connective tissue; Basement membrane;
 KW Extracellular matrix; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DISULFID 1441 1669 NONHELICAL REGION (NC1).
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 FT DISULFID 1505 1511 BY SIMILARITY.
 FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
 FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
 FT DISULFID 1616 1622 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 26 26 A -> P (IN REF. 2).
 FT CONFLICT 186 186 S -> L (IN REF. 2).
 FT CONFLICT 319 319 Q -> S (IN REF. 2).
 FT CONFLICT 359 359 Q -> L (IN REF. 2).
 FT CONFLICT 403 403 L -> P (IN REF. 2).
 FT CONFLICT 481 481 P -> L (IN REF. 2).
 FT CONFLICT 493 493 Q -> H (IN REF. 2).
 FT CONFLICT 712 712 S -> I (IN REF. 2).
 FT CONFLICT 813 813 E -> Q (IN REF. 2).
 FT CONFLICT 982 982 Q -> H (IN REF. 2).
 FT CONFLICT 982 982 V -> S (IN REF. 3).
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).
 SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 72.4%; Score 105; DB 1; Length 1669;
 Best Local Similarity 73.1%; Pred. No. 2.1e-07;
 Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMFLFDNVDNFASENDYS 27

Db 1495 RKFTMPFLFCNNVNCVFASENDYS 1520

RESULT 7

CA24_ASCSU

ID_CA24_ASCSU STANDARD; PRT; 1763 AA.

AC F27353;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.
 OC NCBI_TaxID=6253;
 RN [1]
 RX SEQUENCE FROM N.A. (ISOFORMS I AND II).
 RX MEDLINE=91340768; PubMed=1714907;
 RA Pettitt J., Kingston I.B.;
 RT "The complete primary structure of a nematode alpha 2(IV) collagen
 and the partial structural organization of its gene.";
 RL J. Biol. Chem. 266:16149-16156(1991).
 CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
 CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
 CC Type IV collagen forms a mesh-like network linked through
 CC intermolecular interactions between 7S domains and between NC1
 CC domains.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=I;
 CC IsoId=P27393-1; Sequence=Displayed;
 CC Name=II;
 CC IsoId=P27393-2; Sequence=VSP 001159;
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the
 G-X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M67507; AA018014.1; -
 DR PIR; S16366; S16366.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagnc4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 25.
 DR ProDom; PD000007; C1g_helix; 6.
 DR ProDom; PD003923; Procollagnc4; 1.
 DR SMART; SM00111; C4; 2.
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 KW Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.
 FT DOMAIN 27 42 7S DOMAIN.
 FT DOMAIN 43 1529 TRIPLE-HELICAL REGION.
 FT DOMAIN 1530 1763 NONHELICAL REGION (NC1).
 FT DISULFID 1548 1637 OR 1634 (BY SIMILARITY).
 FT DISULFID 1581 1634 OR 1637 (BY SIMILARITY).
 FT DISULFID 1593 1599 BY SIMILARITY.
 FT DISULFID 1656 1752 OR 1749 (BY SIMILARITY).
 FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).
 FT DISULFID 1702 1709 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 O-LINKED (XYL. . .) (GLYCOSAMINOGLYCAN)
 FT VARSPLIC 230 266 (IN ISOFORM II) (POTENTIAL).
 FT GEQGRGPGPGPGVPSGAKGTIGPSGPMKGEK ->
 FT GDIGPAGPPGPPGPPGPGTIGPSGPMKGEK (in

```
FT isoform II).
FT /FTID=VSP 001159.
SQ SEQUENCE 1763 AA; 168526 MW; 304F52BEC06A80D CRC64;
Query Match 62.8%; Score 91; DB 1; Length 1763;
Best Local Similarity 72.0%; Pred. No. 2.8e-05;
Matches 18; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 3 RFTTWPFLPDVNDVNFASRNDYS 27
||:|||||:|:|:|:|:|:|
Db 1584 RFTTWPFLPDVNDVNFASRNDKS 1608
||:|||||:|:|:|:|:|:|
RESULT 8
ID CA24 CAEEL STANDARD; PRT; 1758 AA.
AC P17140; Q19098; Q19099;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (lethal protein 2).
GN LET-2 OR CUB-1 OR FOIG12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RL J. Cell Biol. 123:253-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RL EMBO J. 13:3278-3285(1994).
RN [4]
RP FUNCTION: Collagen type IV is specific for basement membranes.
CC Vital for embryonic development.
CC -! SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I; Synonyms=a;
CC IsoId=P17140-1; Sequence=Displayed;
CC Name=II; Synonyms=b;
CC IsoId=P17140-2; Sequence=VSP 001160;
CC -! DEVELOPMENTAL STAGE: Isoform I is predominant in embryos and
CC isoform II is predominant in the larvae and adults.
CC -! DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
```

Matches 17; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVDVDFASRNDYS 27

DB 1581 QRFTHMPFLFDNVDVDFASRNDYS 1606

RESULT 9

CA14 CAEEL

ID CA14 CAEEL STANDARD; PRT; 1758 AA.

AC P17139;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 1(IV) chain precursor.

GN EMB-9 OR CLB-2 OR KO4H4.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;

OC Rhabditiidae; Feloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=91141582; PubMed=1996137;

RA Guo X., Johnson J.J., Kramer J.M.;

RT "Embryonic lethality caused by mutations in basement membrane

collagen of C. elegans.";

RL Nature 349:707-709(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,

RA Lattelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,

RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,

RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,

RA Woldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans.";

RL Nature 368:32-38(1994).

RN [3]

RP REVISIONS.

RC Durbin R.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.

RN [4]

RP SEQUENCE OF 1446-1758 FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=9008929; PubMed=2793871;

RA Guo X., Kramer J.M.;

RT "The two Caenorhabditis elegans basement membrane (type IV) collagen

genes are located on separate chromosomes.";

RL Biol. Chem. 264:17574-17582(1989).

CC -1- FUNCTION: Collagen type IV is specific for basement membranes.

CC -1- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.

CC Type IV collagen forms a mesh-like network linked through

intermolecular interactions between 7S domains and through NC1

domains.

CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous

domain (NC1) at their C-terminus, frequent interruptions of the

G-X-Y repeats in the long central triple-helical domain (which may

cause flexibility in the triple helix), and a short N-terminal

triple-helical 7S domain.

CC -1- PTM: Prolines at the third position of the tripeptide repeating

unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -1- PTM: Type IV collagens contain numerous cysteine residues which

are involved in inter- and intramolecular disulfide bonding. 12 of

these, located in the NC1 domain, are conserved in all known type

CC IV collagens.
CC -1- DISEASE: Mutations in this gene cause temperature-sensitive
CC lethality during late embryogenesis.
CC
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X56979; CAA40299.1; -
CC EMBL; Z27078; CAA81584.3; -
CC EMBL; J05067; AAB59179.1; -
CC PIR; S40991; S40991.
CC WormPep; KO4H4.1; C323462.
CC DR InterPro; IPR008161; Clg_helix. 11.
CC DR InterPro; IPR008160; Collagen.
CC DR InterPro; IPR001442; Procollagen4_C.
CC PFam; PF01413; C4; 2.
CC DR Pfam; PF01391; Collagen; 27.
CC DR ProDom; PD000007; Clg_helix; 11.
CC DR ProDom; PD003923; Procollagen4; 1.
CC DR SMART; SM00111; C4; 2.
CC KW Extracellular matrix; Connective tissue; Basement membrane;
CC Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
CC FT SIGNAL 1 20
CC FT PROPEP 21 7194 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
CC FT CHAIN 195 1758 COLLAGEN ALPHA 1(IV) CHAIN.
CC FT DOMAIN 1530 1758 TRIPLE-HELICAL REGION.
CC FT DISULFID 1549 1640 NONHELICAL REGION (NC1).
CC FT DISULFID 1582 1637 OR 1637 (BY SIMILARITY).
CC FT DISULFID 1594 1600 OR 1640 (BY SIMILARITY).
CC FT DISULFID 1659 1754 BY SIMILARITY.
CC FT DISULFID 1693 1751 OR 1751 (BY SIMILARITY).
CC FT DISULFID 1705 1711 OR 1754 (BY SIMILARITY).
CC FT VARIANT 402 402 G -> E (IN MUTANT G34).
CC FT VARIANT 408 408 G -> E (IN MUTANT G23/HCT70).
CC FT CONFLICT 302 391

CC CONFLICT 581 581 G -> R (IN REF. 2).
CC CONFLICT 768 768 P -> R (IN REF. 2).
CC CONFLICT 830 830 D -> V (IN REF. 2).
CC CONFLICT 1514 1514 P -> Q (IN REF. 4).
CC CONFLICT 1722 1722 P -> L (IN REF. 2).
CC SQ SEQUENCE 1758 AA; 170857 MW; 7083D9AF63E05D45 CRC64;

Query Match 57.2%; Score 83; DS 1; Length 1758;

Best Local Similarity 60.0%; Pred. No. 0.00042;

Matches 15; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 RFTMPFLFDNVDVDFASRNDYS 27

DB 1585 KFTMPFLFDNVDVDFASRNDYS 1609

RESULT 10

CA64_HUMAN

ID CA64_HUMAN STANDARD; PRT; 1691 AA.

AC Q14031; Q12823; Q14053; Q9NQM5; Q9NTX3; Q9U076; Q9UMG6; Q9Y4L4;

DT 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Collagen alpha 6(IV) chain precursor.

GN COL4A6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eye, and Kidney;
RX MEDLINE=94171779; PubMed=8125972;
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.,
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RL J. Biol. Chem. 269:7520-7526(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Reders S.T.,
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Reders S.T., Tryggvason K.,
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grahame D., Lawlor S., Wilson S.,
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=8356443;
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,
RA de Paeppe A., Tryggvason K., Reders S.T.,
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169(1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D21337; BAA04809.1; -
DR EMBL; U04845; AAA19569.2; -
DR EMBL; U47004; AAB19038.1; -
DR EMBL; U46959; AAB19038.1; JOINED.
DR EMBL; U46961; AAB19038.1; JOINED.
DR EMBL; U46962; AAB19038.1; JOINED.
DR EMBL; U46963; AAB19038.1; JOINED.
DR EMBL; U46964; AAB19038.1; JOINED.
DR EMBL; U46965; AAB19038.1; JOINED.
DR EMBL; U46966; AAB19038.1; JOINED.
DR EMBL; U46967; AAB19038.1; JOINED.
DR EMBL; U46968; AAB19038.1; JOINED.
DR EMBL; U46969; AAB19038.1; JOINED.
DR EMBL; U46970; AAB19038.1; JOINED.
DR EMBL; U46971; AAB19038.1; JOINED.
DR EMBL; U46972; AAB19038.1; JOINED.
DR EMBL; U46973; AAB19038.1; JOINED.
DR EMBL; U46974; AAB19038.1; JOINED.
DR EMBL; U46975; AAB19038.1; JOINED.
DR EMBL; U46976; AAB19038.1; JOINED.
DR EMBL; U46977; AAB19038.1; JOINED.
DR EMBL; U46978; AAB19038.1; JOINED.
DR EMBL; U46979; AAB19038.1; JOINED.
DR EMBL; U46980; AAB19038.1; JOINED.
DR EMBL; U46981; AAB19038.1; JOINED.
DR EMBL; U46982; AAB19038.1; JOINED.
DR EMBL; U46983; AAB19038.1; JOINED.
DR EMBL; U46984; AAB19038.1; JOINED.
DR EMBL; U46985; AAB19038.1; JOINED.
DR EMBL; U46986; AAB19038.1; JOINED.
DR EMBL; U46987; AAB19038.1; JOINED.
DR EMBL; U46988; AAB19038.1; JOINED.
DR EMBL; U46989; AAB19038.1; JOINED.
DR EMBL; U46990; AAB19038.1; JOINED.
DR EMBL; U46991; AAB19038.1; JOINED.
DR EMBL; U46992; AAB19038.1; JOINED.
DR EMBL; U46993; AAB19038.1; JOINED.
DR EMBL; U46994; AAB19038.1; JOINED.
DR EMBL; U46995; AAB19038.1; JOINED.
DR EMBL; U46996; AAB19038.1; JOINED.
DR EMBL; U46997; AAB19038.1; JOINED.
DR EMBL; U46998; AAB19038.1; JOINED.
DR EMBL; U46999; AAB19038.1; JOINED.
DR EMBL; U47000; AAB19038.1; JOINED.
DR EMBL; U47001; AAB19038.1; JOINED.
DR EMBL; U47002; AAB19038.1; JOINED.
DR EMBL; U47003; AAB19038.1; JOINED.
DR EMBL; U47004; AAB19039.1; -
DR EMBL; U46960; AAB19039.1; JOINED.
DR EMBL; U46961; AAB19039.1; JOINED.
DR EMBL; U46962; AAB19039.1; JOINED.
DR EMBL; U46963; AAB19039.1; JOINED.
DR EMBL; U46964; AAB19039.1; JOINED.
DR EMBL; U46965; AAB19039.1; JOINED.
DR EMBL; U46966; AAB19039.1; JOINED.
DR EMBL; U46967; AAB19039.1; JOINED.
DR EMBL; U46968; AAB19039.1; JOINED.
DR EMBL; U46969; AAB19039.1; JOINED.
DR EMBL; U46970; AAB19039.1; JOINED.
DR EMBL; U46971; AAB19039.1; JOINED.
DR EMBL; U46972; AAB19039.1; JOINED.
DR EMBL; U46973; AAB19039.1; JOINED.
DR EMBL; U46974; AAB19039.1; JOINED.
DR EMBL; U46975; AAB19039.1; JOINED.
DR EMBL; U46976; AAB19039.1; JOINED.
DR EMBL; U46977; AAB19039.1; JOINED.
DR EMBL; U46978; AAB19039.1; JOINED.
DR EMBL; U46979; AAB19039.1; JOINED.
DR EMBL; U46980; AAB19039.1; JOINED.
DR EMBL; U46981; AAB19039.1; JOINED.
DR EMBL; U46982; AAB19039.1; JOINED.
DR EMBL; U46983; AAB19039.1; JOINED.

```

DR EMBL; U46984; AAB19039.1; JOINED.
DR EMBL; U46985; AAB19039.1; JOINED.
DR EMBL; U46986; AAB19039.1; JOINED.
DR EMBL; U46987; AAB19039.1; JOINED.
DR EMBL; U46988; AAB19039.1; JOINED.
DR EMBL; U46989; AAB19039.1; JOINED.
DR EMBL; U46990; AAB19039.1; JOINED.
DR EMBL; U46991; AAB19039.1; JOINED.
DR EMBL; U46992; AAB19039.1; JOINED.
DR EMBL; U46993; AAB19039.1; JOINED.
DR EMBL; U46994; AAB19039.1; JOINED.
DR EMBL; U46995; AAB19039.1; JOINED.
DR EMBL; U46996; AAB19039.1; JOINED.
DR EMBL; U46997; AAB19039.1; JOINED.
DR EMBL; U46998; AAB19039.1; JOINED.
DR EMBL; U46999; AAB19039.1; JOINED.
DR EMBL; U47000; AAB19039.1; JOINED.
DR EMBL; U47001; AAB19039.1; JOINED.
DR EMBL; U47002; AAB19039.1; JOINED.
DR EMBL; U47003; AAB19039.1; JOINED.
DR EMBL; AL034369; CAA22265.1; -.
DR EMBL; AL109943; CAB89263.1; -.
DR EMBL; AL136085; CAB96748.1; -.
DR EMBL; AL031177; CAA20120.1; -.
DR EMBL; L22763; AAA16338.1; -.
DR PIR; A54122; CGHU6B.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR GO; GO:0005201; E:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; P:extracellular matrix organization and biogen. .; NAS.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; C1g_helix; 4.
DR SMART; SM00111; C4; 2.
DR SMART; SM00111; C4; 2.
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.

Query Match 55.9%; Score 81; DB 1; Length 1691;
Best Local Similarity 56.0%; Pred. No. 0.00081;
Matches 14; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 3 RFTTWPFLEFDVNVDFASRNDYS 27
DB 1518 RFTTWPFLEFDVNVDFASRNDYS 1542
|||||:::|:::|:::|:::|
|||||:::|:::|:::|:::|

RESULT 11
ID CA24_MOUSE STANDARD; PRT; 1707 AA.
AC P08122; Q61375;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX MEDLINE=8917933; PubMed=2703491;
RP SEQUENCE FROM N.A.
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumaran G.,
RA Pihlajaniemi T., Kurkinen M.;

```

```

RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen."
RL J. Biol. Chem. 264:6318-6324(1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RT Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes."
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
RA Deutzmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RT chain and its comparison with the alpha 1(IV) chain."
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1(IV)."
RL FEBS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen."
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen."
RL FEBS Lett. 206:29-32(1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RT alpha 2(IV) collagen gene."
RL Nature 317:177-179(1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdalo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer."
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.

```


RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RL of basement-membrane type IV collagen.";
 CC Eur. J. Biochem. 176:617-624 (1988).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire',
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 CC alpha 6(IV), each of which can form a triple helix structure
 CC with 2 other chains to generate type IV collagen network.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the
 CC G-X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; X05562; CAA29078.1; -
 CC EMBL; X05610; CAA29098.1; -
 CC EMBL; J02760; AAB58422.1; -
 CC EMBL; M36963; AAB53099.1; -
 CC EMBL; X12784; CAA31275.1; -
 CC EMBL; J04217; AAB53097.1; -
 CC PIR; A32024; CCHU2B.
 CC Genew; HGNC:2203; COL4A2.
 CC MIM; 120090; -
 CC GO; GO:0005587; C:collagen type IV; TAS.
 CC GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
 CC GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
 CC InterPro; IPR008161; Clg.helix.
 CC InterPro; IPR008160; Collagen.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 24.
 CC ProDom; PD000007; Clg.helix; 7.
 CC ProDom; PD003923; ProcollagnC4; 1.
 CC SMART; SM00111; C4; 2.
 CC XW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC KW Glycoprotein; Basement membrane; Collagen; Signal.
 CC FT SIGNAL 1 25
 CC FT PROPEP 16 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 CC FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
 CC FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.
 CC FT DISULFID 1485 1712 NONHELICAL REGION (NC1).
 CC FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
 CC FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
 CC FT DISULFID 1549 1555 BY SIMILARITY.
 CC FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
 CC FT DISULFID 1645 1705 OR 1708 (BY SIMILARITY).
 CC FT DISULFID 1658 1665 BY SIMILARITY.
 CC FT CARBOHYD 138 138 N-LINKED (GLCNAC. .).
 CC FT CONFLICT 471 471 R -> P (IN REF. 2).
 CC FT CONFLICT 683 683 A -> G (IN REF. 2).
 CC FT CONFLICT 1575 1575 M -> I (IN REF. 5).
 CC FT CONFLICT 1663 1663 G -> H (IN REF. 9).
 CC FT CONFLICT 1701 1701 H -> G (IN REF. 9).
 CC SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
 CC Query Match 53.8%; Score 78; DB 1; Length 1712;
 CC Best Local Similarity 64.0%; Pred. No. 0.0023;

Matches 16; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 3 RFTTTPFLFDVNDVNDVNFASRNDYS 27
 Db 1540 RSTTTPFLYCNPGDVCYVSRNDKS 1564
 ||:||||:| ||:|||||
 1540 RSTTTPFLYCNPGDVCYVSRNDKS 1564
 RESULT 13
 CA44_RABIT
 ID CA44_RABIT STANDARD; PRT; 623 AA.
 AC P55787;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Collagen alpha 4(IV) chain (Fragment).
 GN COL4A4.
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CC NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Corneal endothelium;
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 RT alpha 4 chain of basement membrane collagen type IV and assignment of
 RT the gene to the distal long arm of human Chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758 (1992).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire',
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 CC alpha 6(IV), each of which can form a triple helix structure with
 CC 2 other chains to generate type IV collagen network.
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-
 CC X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; L01477; -; NOT_ANNOTATED_CDS.
 CC PIR; A45137; A45137.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR001442; Procollagn4_C.
 CC Pfam; PF01413; C4; 2.
 CC Pfam; PF01391; Collagen; 5.
 CC ProDom; PD003923; ProcollagnC4; 1.
 CC SMART; SM00111; C4; 2.
 CC XW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 CC NON_TER 1 1
 CC DOMAIN <1 392 TRIPLE-HELICAL REGION.
 CC FT DOMAIN 393 623 NONHELICAL REGION (NC1).
 CC FT DISULFID 413 502 OR 499 (BY SIMILARITY).
 CC FT DISULFID 446 499 OR 502 (BY SIMILARITY).


```
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD003923; Procollagen4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38
FT CHAIN 39 1690
FT DOMAIN 39 64
FT DOMAIN 65 1459
FT DOMAIN 1460 1690
FT SITE 94 96
FT SITE 145 147
FT SITE 189 191
FT SITE 310 312
FT SITE 724 726
FT SITE 785 787
FT SITE 989 991
FT SITE 1206 1207
FT SITE 1212 1214
FT DISULFID 1480 1569
FT DISULFID 1513 1566
FT DISULFID 1525 1531
FT DISULFID 1588 1686
FT DISULFID 1622 1683
FT DISULFID 1634 1641
FT CARBOHYD 142 142
FT CARBOHYD 669 669
FT VARIANT 441 446
FT VARIANT 545 545
FT VARIANT 570 570
FT VARIANT 897 897
FT VARIANT 931 931
FT VARIANT 1004 1004
FT VARIANT 1030 1030
FT VARIANT 1201 1201
FT VARIANT 1402 1402
FT VARIANT 1572 1572
FT CONFLICT 1659 1660
FT SEQUENCE 1690 AA; 164095 MW; E1E72F283A72BAAE CRC64;
Query Match 44.1%; Score 64; DB 1; Length 1690;
Best local Similarity 45.8%; Pred. No. 0.27; Mismatches 6; Indels 0; Gaps 0;
Matches 11; Conservative 7;
QY 4 FTTPFPFLPDVNDVDFASRNDYS 27
Db 1517 FSTLPFAFCNIHQVCHYQNRDS 1540
RESULT 15
ID CA14_DROME STANDARD; PRT; 1775 AA.
AC P08120;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
```

DR EMBL; M28334; AAA28422.1; -;
DR EMBL; V00200; CAA23486.2; -;
DR PIR; A31893; A31893;
DR Flybase; FBgn0000299; Cg25C.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2;
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD000007; C1g_helix; 9.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 ?
FT CHAIN ? 1775 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT DOMAIN ? 1544 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 1545 1775 TRIPLE-HELICAL REGION.
FT DISULFID 1569 1655 NONHELICAL REGION (NC1).
FT DISULFID 1599 1652 OR 1652 (BY SIMILARITY).
FT DISULFID 1811 1817 OR 1655 (BY SIMILARITY).
FT DISULFID 1874 1770 BY SIMILARITY.
FT DISULFID 1708 1767 OR 1767 (BY SIMILARITY).
FT DISULFID 1720 1727 OR 1770 (BY SIMILARITY).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 948 948 L -> S (IN REF. 6).
FT CONFLICT 997 997 S -> T (IN REF. 6).
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).
FT CONFLICT 1373 1373 T -> I (IN REF. 5).
FT CONFLICT 1496 1496 L -> R (IN REF. 5).
FT CONFLICT 1507 1511 ETGNV -> RAGOR (IN REF. 5).
FT CONFLICT 1529 1529 E -> K (IN REF. 5).
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DESAB23149525CD CRC64;

Query Match 44.1%; Score 64; DB 1; Length 1775;
Best Local Similarity 56.5%; Pred. No. 0.29;
Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 3 RFTTPEFLFDVNDVDFASRND 25
|:|:| |:|:|:|
Db 1602 RFTLPVLSCQNNVCNYSRND 1624

Search completed: April 5, 2004, 06:59:42
Job time : 3.39952 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 16.2131 seconds
(without alignments)
525.440 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTTPFLFDVNDVDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 122 | 84.1 | 212 | 6 Q28512 | Q28512 macaca mula |
| 2 | 122 | 84.1 | 245 | 4 Q9NYC4 | Q9NYC4 homo sapien |
| 3 | 121 | 83.4 | 203 | 6 Q29032 | Q29032 sus scrofa |
| 4 | 121 | 83.4 | 203 | 6 Q28682 | Q28682 oryctolagus |
| 5 | 121 | 83.4 | 212 | 6 Q28567 | Q28567 ovis aries |
| 6 | 116 | 80.0 | 161 | 11 Q61430 | Q61430 mus musculus |
| 7 | 116 | 80.0 | 210 | 6 Q28273 | Q28273 canis famil |
| 8 | 116 | 80.0 | 246 | 11 Q61435 | Q61435 mus musculus |
| 9 | 116 | 80.0 | 1669 | 11 Q9Q2S0 | Q9Q2S0 mus musculus |
| 10 | 112 | 77.2 | 230 | 11 Q63122 | Q63122 rattus norv |
| 11 | 106 | 73.1 | 179 | 11 P70165 | P70165 mus musculus |
| 12 | 106 | 73.1 | 253 | 11 Q61436 | Q61436 mus musculus |
| 13 | 106 | 73.1 | 585 | 11 Q80V57 | Q80V57 mus musculus |
| 14 | 106 | 73.1 | 799 | 11 Q8BNS7 | Q8BNS7 mus musculus |
| 15 | 106 | 73.1 | 886 | 4 Q9NUS7 | Q9NUS7 homo sapien |
| 16 | 106 | 73.1 | 1684 | 6 Q8HYC1 | Q8HYC1 canis famil |

| | | | | | |
|----|-----|------|------|-----------|---------------------|
| 17 | 106 | 73.1 | 1688 | 6 Q86G22 | Q86G22 canis famil |
| 18 | 106 | 73.1 | 1691 | 11 Q9BSQ2 | Q9BSQ2 mus musculus |
| 19 | 105 | 72.4 | 226 | 6 Q28271 | Q28271 canis famil |
| 20 | 105 | 72.4 | 226 | 11 Q9SLQ8 | Q9SLQ8 mus musculus |
| 21 | 105 | 72.4 | 229 | 4 Q8NFB8 | Q8NFB8 homo sapien |
| 22 | 105 | 72.4 | 229 | 4 Q9NYC5 | Q9NYC5 homo sapien |
| 23 | 105 | 72.4 | 979 | 13 Q919K3 | Q919K3 gallus gall |
| 24 | 105 | 72.4 | 1075 | 4 Q86X41 | Q86X41 homo sapien |
| 25 | 105 | 72.4 | 1621 | 4 Q9H4R9 | Q9H4R9 homo sapien |
| 26 | 96 | 66.2 | 1747 | 5 Q26640 | Q26640 strongyloce |
| 27 | 96 | 66.2 | 1752 | 5 Q27265 | Q27265 strongyloce |
| 28 | 84 | 57.9 | 1802 | 5 Q17163 | Q17163 brugia mala |
| 29 | 81 | 55.9 | 205 | 6 Q28274 | Q28274 canis famil |
| 30 | 81 | 55.9 | 546 | 11 Q99K97 | Q99K97 mus musculus |
| 31 | 81 | 55.9 | 1600 | 4 Q9UEH6 | Q9UEH6 homo sapien |
| 32 | 81 | 55.9 | 1691 | 11 Q9BSQ1 | Q9BSQ1 mus musculus |
| 33 | 78 | 53.8 | 202 | 6 Q28272 | Q28272 canis famil |
| 34 | 78 | 53.8 | 358 | 11 Q91V13 | Q91V13 mus musculus |
| 35 | 78 | 53.8 | 673 | 4 Q14052 | Q14052 homo sapien |
| 36 | 75 | 51.7 | 1723 | 5 Q9GQ31 | Q9GQ31 hydra atten |
| 37 | 65 | 44.8 | 1761 | 5 Q18407 | Q18407 drosophila |
| 38 | 65 | 44.8 | 1940 | 5 Q9VMV5 | Q9VMV5 drosophila |
| 39 | 64 | 44.1 | 312 | 11 Q64457 | Q64457 mus musculus |
| 40 | 64 | 44.1 | 1682 | 11 Q9QZ99 | Q9QZ99 mus musculus |
| 41 | 64 | 44.1 | 1779 | 5 Q9VMV4 | Q9VMV4 drosophila |
| 42 | 63 | 43.4 | 208 | 6 Q29468 | Q29468 canis famil |
| 43 | 63 | 43.4 | 1024 | 5 Q8T7S4 | Q8T7S4 anopheles g |
| 44 | 60 | 41.4 | 713 | 5 Q9GV24 | Q9GV24 sarcophaga |
| 45 | 59 | 40.7 | 371 | 8 Q8SGX2 | Q8SGX2 calanaria p |

ALIGNMENTS

RESULT 1

ID Q28512 PRELIMINARY; PRT; 212 AA.

AC Q28512; 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Alpha-3 type IV collagen (Fragment).

GN COL4A3.

OS Macaca mulatta (Rhesus macaque).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

OC Cercopithecoidea; Macaca.

OC NCBI_TaxID=9544;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney cortex;

RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,

Mason P.J., Pusey C.D.;

RT "Properties and sequences of the Goodpasture antigen of different

mammals.;"

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; L47380; AA91861.1; -

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollag4.C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; Procollag4; 1.

DR SMART; SM00111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

DR Collagen.

FT NON_TER 1

FT NON_TER 212

SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;

Query Match 84.1%; Score 122; DB 6; Length 212;

Best Local Similarity 92.3%; Pred. No. 1.8e-10;


```

DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
FT NON_TER 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 83.4%; Score 121; DB 6; Length 203;
Best Local Similarity 88.5%; Pred.No.2.4e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFDNVNDVNFASRNDYS 27
      ||||| : ||| |||||
DB 37 QRFTTMPFLFCNVDCVNFASRNDYS 62
      ||||| : ||| |||||

RESULT 4
Q28682 PRELIMINARY; PRT; 203 AA.
ID Q28682 AC Q28682
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_TaxID=9986;
[1] SEQUENCE FROM N.A.
RP TISSUE=Kidney cortex;
RC Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; I47283; AAA91893.1; ".
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E1417381E4D9E30 CRC64;

Query Match 83.4%; Score 121; DB 6; Length 203;
Best Local Similarity 88.5%; Pred.No.2.4e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFDNVNDVNFASRNDYS 27
      ||||| : ||| |||||
DB 37 QRFTTMPFLFCNVDCVNFASRNDYS 62
      ||||| : ||| |||||

RESULT 5
Q28567 PRELIMINARY; PRT; 212 AA.
ID Q28567 AC Q28567
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).

```

| | | | | |
|----|-----------|-------------|--|-------------------------|
| DR | GO: | GO:0005581; | C:collagen; | IEA. |
| DR | GO: | GO:0005201; | F:extracellular matrix structural constituent; | IEA. |
| DR | GO: | GO:003676; | F:nucleic acid binding; | IEA. |
| DR | InterPro: | IPR001442; | Prscollagn4_C. | |
| DR | InterPro: | IPR000504; | RNA_rec_mot. | |
| DR | Pfam: | PF01413; | C4; 2. | |
| DR | SMART: | SM00111; | C4; 2. | |
| DR | PROSITE: | PS00030; | RRM_RNP_1; 1. | |
| FT | NON_TER | 1 | I | |
| FT | NON_TER | 161 | 161 | |
| SQ | SEQUENCE | 161 AA; | 17925 MW; | 1F59DF6CFE8236C5 CRC64; |

| | |
|-------|--|
| COLR3 | Mus musculus (Mouse). |
| GN | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; |
| OX | NCBI_TaxID=10090; |
| | [1]_TaxID=10090; |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=Balb/C; |
| RX | MEDLINE=95050957; PubMed=7962065; |
| RA | Miner J.H., Sanes J.R.; |
| RT | "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal |

RT laminae: Sequence, distribution, association with laminins, and
 RT developmental switches."
 RL J. Cell Biol. 127:879-891(1994).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Balb/c;
 RA Miner J.H.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 DR EXBL; Z35166; CAA84529.1; -.

DR PIR; I48302; I48302.

DR MGD; MGI:104688; Col4a3.

DR GO; GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR001442; Procollagn4_C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SM00111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

FT NON_TER 1

SQ SEQUENCE 246 AA; 26993 MW; A9B5434P5836F324 CRC64;

Query Match 80.0%; Score 116; DB 11; Length 246;

Best Local Similarity 84.6%; Pred. No. 1.7e-09;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRTTTPFLFDNVNDVNFASRNDYS 27

DB 71 QRTTTPFLFCNNVNCVFASRNDYS 96

RESULT 9

O9QZS0

ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.

AC Q9QZS0;

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Alpha 3 collagen IV.

GN COL4A3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ON NCBI_TaxID=10090;

OX [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=20005934; PubMed=10534397;

RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,

RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;

RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a

RT mouse model of alport syndrome.";

RL Genomics 61:113-124(1999).

DR EMBL; AF169387; AAD50449.1; -.

DR PIR; I48302; I48302.

DR MGD; MGI:104688; Col4a3.

DR GO; GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR008161; C1g_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR001442; Procollagn4_C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 21.

DR ProDom; PD000007; C1g_helix; 6.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SM00111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

KW Collagen.

SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A7B2 CRC64;

Query Match

Best Local Similarity 80.0%; Score 116; DB 11; Length 1669;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRTTTPFLFDNVNDVNFASRNDYS 27

DB 1494 QRTTTPFLFCNNVNCVFASRNDYS 1519

RESULT 10

Q63122

ID Q63122 PRELIMINARY; PRT; 230 AA.

AC Q63122;

DT 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Alpha-3 type IV collagen (Fragment).

GN COL4A3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

ON NCBI_TaxID=10116;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

RX MEDLINE=98210005; PubMed=950634;

RA Ryan J.J., Katbanna I., Mason P.J., Pusey C.D., Turner A.N.;

RT "Sequence analysis of the 'Goodpasture antigen' of mammals."

RL Nephrol. Dial. Transplant. 13:602-607(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

RA Turner N.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; I47281; AAB72338.2; -.

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollagn4_C.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SM00111; C4; 2.

DR PROSITE; PS00030; RRM_RNP_1; 1.

KW Collagen.

FT NON_TER 1

FT NON_TER 230

SQ SEQUENCE 230 AA; 25398 MW; 29549E2514CC056 CRC64;

Query Match

Best Local Similarity 77.2%; Score 112; DB 11; Length 230;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRTTTPFLFDNVNDVNFASRNDYS 27

DB 55 QRTTTPFLFCNNVNCVFASRNDYS 80

RESULT 11

F70165

ID F70165 PRELIMINARY; PRT; 179 AA.

AC F70165;

DT 01-FEB-1997 (TRENBLrel. 02, Created)

DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Collagen type IV alpha5 chain (Fragment).

GN COL4A5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ON NCBI_TaxID=10090;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129;

RT Oberbauer I.;

RT "Cloning of the NCI domains of the minor collagen IV chains of mouse

RT via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and

```

RT  alphas(IV) in differentiated teratocarcinoma cells.";
RL  Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR  EMBL; X82218; CAA57698.1; -.
DR  GO; GO:0005931; C:collagen; IEA.
DR  GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR  InterPro; IPR001442; Procollagn4_C.
DR  Pfam; PF01413; C4; 2.
DR  ProDom; PD003923; ProcollagnC4; 1.
DR  SMART; SM00111; C4; 2.
FT  NON_TER 179 179
FT  NON_TER 179 179
SQ  SEQUENCE 179 AA; 13959 MW; 20A188F3687F582F CRC64;

Query Match 73.1%; Score 106; DB 11; Length 179;
Best Local Similarity 73.1%; Pred. No. 4.1e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY  2 QRFTHMPFLFDNVNDVNFASRNDYS 27
    :||:||||:|:|:|||||
Db  32 RRFSTMPFMCNINNVCFASRNDYS 57

RESULT 12
Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: Sequence, distribution, association with laminins, and
RT developmental switches.";
RL J Cell Biol. 127:879-891(1994).
DR EMBL; Z35168; CAA84531.1; -.
DR FIR; I48304; I48304.
DR MGD; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER 1 1
FT NON_TER 1 1
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match 73.1%; Score 106; DB 11; Length 253;
Best Local Similarity 73.1%; Pred. No. 6e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY  2 QRFTHMPFLFDNVNDVNFASRNDYS 27
    :||:||||:|:|:|||||
Db  79 RRFSTMPFMCNINNVCFASRNDYS 104

RESULT 13
Q80V57 PRELIMINARY; PRT; 585 AA.
AC Q80V57;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Col4a5 protein.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Shat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M.J., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043317; AAH43317.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
SQ SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match 73.1%; Score 106; DB 11; Length 585;
Best Local Similarity 73.1%; Pred. No. 1.5e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY  2 QRFTHMPFLFDNVNDVNFASRNDYS 27
    :||:||||:|:|:|||||
Db  411 RRFSTMPFMCNINNVCFASRNDYS 436

RESULT 14
Q8BNS7 PRELIMINARY; PRT; 799 AA.
AC Q8BNS7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354583; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```

```
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGI; 88456; Col14a5.
GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 799 AA; 7789 MW; C517CF4CF15705DC CRC64;

Query Match 73.1%; Score 106; DB 11; Length 799;
Best Local Similarity 73.1%; Pred. No. 2.4e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDNFASRNDYS 27
DB 625 RRFSTMPFNFNCNNVCFASRNDYS 650

RESULT 15
Q9NUE7
ID Q9NUE7 PRELIMINARY; PRT; 886 AA.
AC Q9NUE7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RA Cobley V.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 73.1%; Score 106; DB 4; Length 886;
Best Local Similarity 73.1%; Pred. No. 2.4e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDNFASRNDYS 27
DB 712 RRFSTMPFNFNCNNVCFASRNDYS 737

Search completed: April 5, 2004, 07:03:59
Job time : 17.2131 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 24.3196 Seconds
(without alignments)
313.688 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTTPFLFDVNDVNFASRDYD 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1980s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2000s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 145 | 100.0 | 27 | 6 ADA20241 | Ada20241 P2 peptid |
| 2 | 133 | 91.7 | 27 | 6 ADA20239 | Ada20239 T8-3 pept |
| 3 | 127 | 87.6 | 27 | 6 ADA20238 | Ada20238 T8 peptid |
| 4 | 122 | 84.1 | 79 | 5 AAU75600 | AAU75600 Human typ |
| 5 | 122 | 84.1 | 79 | 6 ADA20264 | Ada20264 Human tum |
| 6 | 122 | 84.1 | 88 | 5 AAU75608 | AAU75608 Human typ |
| 7 | 122 | 84.1 | 88 | 5 AAU75607 | AAU75607 Human typ |
| 8 | 122 | 84.1 | 88 | 6 ADA20271 | Ada20271 Human tum |
| 9 | 122 | 84.1 | 88 | 6 ADA20272 | Ada20272 Human tum |
| 10 | 122 | 84.1 | 124 | 5 AAU75594 | AAU75594 Human typ |
| 11 | 122 | 84.1 | 124 | 6 ADA20258 | Ada20258 Human tum |
| 12 | 122 | 84.1 | 132 | 5 AAU75597 | AAU75597 Human typ |
| 13 | 122 | 84.1 | 132 | 6 ADA20261 | Ada20261 Human tum |
| 14 | 122 | 84.1 | 191 | 5 AAU75596 | AAU75596 Human typ |
| 15 | 122 | 84.1 | 191 | 6 ADA20260 | Ada20260 Human tum |
| 16 | 122 | 84.1 | 211 | 3 AAU755918 | AAU755918 Human Goo |
| 17 | 122 | 84.1 | 211 | 5 ABG79208 | ABG79208 Human GP |
| 18 | 122 | 84.1 | 218 | 2 AAU79164 | AAU79164 Partial s |
| 19 | 122 | 84.1 | 218 | 2 AAU44172 | AAU44172 Human typ |
| 20 | 122 | 84.1 | 218 | 3 AAU56784 | AAU56784 Human alp |
| 21 | 122 | 84.1 | 218 | 4 AAU09484 | AAU09484 Human alp |
| 22 | 122 | 84.1 | 232 | 7 ADC17697 | ADC17697 Human typ |
| 23 | 122 | 84.1 | 244 | 5 ABG79218 | ABG79218 Human typ |
| 24 | 122 | 84.1 | 244 | 5 ABG79219 | ABG79219 Human Goo |
| 25 | 122 | 84.1 | 244 | 5 ABG79217 | ABG79217 Human typ |

| | | | | | |
|----|-----|------|------|------------|--------------------|
| 26 | 122 | 84.1 | 244 | 5 AAU75595 | AAU75595 Human typ |
| 27 | 122 | 84.1 | 244 | 6 ADA20225 | Ada20225 Human typ |
| 28 | 122 | 84.1 | 245 | 3 AAU67942 | AAU67942 Human typ |
| 29 | 122 | 84.1 | 245 | 5 AAU75589 | AAU75589 Human typ |
| 30 | 122 | 84.1 | 254 | 5 AAU75598 | AAU75598 Human typ |
| 31 | 122 | 84.1 | 268 | 2 AAU31993 | AAU31993 Type IV c |
| 32 | 122 | 84.1 | 268 | 3 AAU97555 | AAU97555 Human alp |
| 33 | 122 | 84.1 | 1670 | 7 ADD47063 | Add47063 Human pro |
| 34 | 121 | 83.4 | 471 | 2 AAU79163 | AAU79163 Partial s |
| 35 | 121 | 83.4 | 471 | 2 AAU44171 | AAU44171 Bovine ty |
| 36 | 121 | 83.4 | 471 | 3 AAU56783 | AAU56783 Bovine al |
| 37 | 121 | 83.4 | 471 | 4 AAU09483 | AAU09483 Bovine al |
| 38 | 112 | 77.2 | 230 | 7 ADD47061 | Add47061 Rat Prote |
| 39 | 106 | 73.1 | 229 | 7 ADC17699 | ADC17699 Human typ |
| 40 | 106 | 73.1 | 264 | 2 AAU31995 | AAU31995 Type IV c |
| 41 | 106 | 73.1 | 264 | 3 AAU97557 | AAU97557 Human alp |
| 42 | 106 | 73.1 | 309 | 3 AAU54044 | AAU54044 Human pan |
| 43 | 106 | 73.1 | 772 | 2 AAU23873 | AAU23873 Human alp |
| 44 | 106 | 73.1 | 772 | 2 AAU09643 | AAU09643 Human typ |
| 45 | 106 | 73.1 | 1685 | 4 ABG04839 | ABG04839 Novel hum |

ALIGNMENTS

RESULT 1

ADA20241

ID ADA20241 standard; peptide; 27 AA.

AC ADA20241;

XX 20-NOV-2003 (first entry)

XX P2 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX metastasis; basement membrane organisation; type IV collagen network;

XX C-terminal globular non-collagenous domain; NCI; type IV collagen;

XX cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX cytosolic; gene therapy; P2 peptide; tumstatin; human;

XX type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

FT Misc-difference 12 /note= "Wild-type Cys substituted by Asp"

FT Misc-difference 18 /note= "Wild-type Cys substituted by Asp"

WO2003059257-A2.

24-JUL-2003.

20-DEC-2002; 2002WO-US040938.

21-DEC-2001; 2001US-00032221.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

New peptide, useful for preparing a composition for inhibiting tumor

growth, angiogenic activity or protein synthesis in a mammalian tissue.

Claim 65; Page 45; 240pp; English.

This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the P2 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 CC
 XX Sequence 27 AA;

Query Match 100.0%; Score 145; DB 6; Length 27;
 Best Local Similarity 100.0%; Pred. No. 2.9e-16;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFDNVNVDNFAASNDYS 27
 DB 1 KQRTTTPFLFDNVNVDNFAASNDYS 27

RESULT 2
 ADA20239
 ID ADA20239 standard; peptide; 27 AA.

XX ADA20239;
 XX 20-NOV-2003 (first entry)
 XX T8-3 peptide related to human type IV collagen alpha and angiogenesis.
 XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 XX metastasis; basement membrane organisation; type IV collagen network;
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
 XX cytostatic; gene therapy; T8-3 peptide; tumstatin; human;
 XX type IV collagen alpha 3 chain; mutant; mutein.

OS Synthetic.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"
 FT Misc-difference 12 /note= "Wild-type Cys substituted by Ser"
 FT Misc-difference 18 /note= "Wild-type Cys substituted by Ser"

XX WO2003059257-A2.
 XX 24-JUL-2003.
 XX 20-DEC-2002; 2002WO-US040938.
 XX 21-DEC-2001; 2001US-00032221.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX Kalluri R;
 XX WPI; 2003-587256/55.

PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 PS Claim 63; Page 45; 240pp; English.
 XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the T8-3 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 XX
 XX Sequence 27 AA;

Query Match 91.7%; Score 133; DB 6; Length 27;
 Best Local Similarity 92.6%; Pred. No. 2.6e-14;
 Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFDNVNVDNFAASNDYS 27
 DB 1 KQRTTTPFLFDNVNVDNFAASNDYS 27

RESULT 3
 ADA20238
 ID ADA20238 standard; peptide; 27 AA.

XX ADA20238;
 XX 20-NOV-2003 (first entry)
 XX T8 peptide related to human type IV collagen alpha and angiogenesis.
 XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 XX metastasis; basement membrane organisation; type IV collagen network;
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
 XX cytostatic; gene therapy; T8 peptide; tumstatin; human;
 XX type IV collagen alpha 3 chain; mutant; mutein.

OS Synthetic.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"
 FT WO2003059257-A2.
 XX 24-JUL-2003.
 XX 20-DEC-2002; 2002WO-US040938.
 XX 21-DEC-2001; 2001US-00032221.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;
 XX

DR WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX

XX Claim 62; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries

CC from pre-existing vessels is essential for tumor growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-

CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular

CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

CC collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cytostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present

CC sequence is the amino acid sequence of the T8 peptide of the invention,

CC derived from the amino acid sequence of tumstatin, which in turn was

CC derived from the amino acid sequence of human type IV collagen alpha 3

CC chain.

XX

XX Sequence 27 AA;

SQ

Query Match 87.6%; Score 127; DB 6; Length 27;

Best Local Similarity 92.6%; Pred. No. 2.4e-13;

Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFDNVDVNFASRNDYS 27

DB 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

RESULT 4

AAU75600

ID AAU75600 standard; protein; 79 AA.

XX

AC AAU75600;

XX

XX 08-MAY-2002 (first entry)

DT

XX

XX Human type IV collagen alpha 3 chain mutant, Tum-5.

DE

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;

KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX

OS Homo sapiens.

XX

XX WO200151523-A2.

PN

XX

XX 19-JUL-2001.

PD

XX

XX 08-JAN-2001; 2001WO-US000565.

PF

XX

XX 07-JAN-2000; 2000US-00479118.

PR

XX

XX 04-APR-2000; 2000US-00543371.

PR

XX

XX 21-JUL-2000; 2000US-00625191.

PR

XX

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA

XX

XX Kalluri R;

PI

XX

XX WPI; 2002-188037/24.

DR

XX

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

PT

treating disorders involving angiogenesis.

Example 40; Page: 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues 54-132 of Tumstatin. Note: The present sequence is not shown in the specification but is derived from the wild type human Tumstatin sequence given in figure 18A (see AAU75589)

XX

XX Sequence 79 AA;

SQ

Query Match 84.1%; Score 122; DB 5; Length 79;

Best Local Similarity 92.3%; Pred. No. 5.9e-12;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFDNVDVNFASRNDYS 27

DB 17 QRFTTTPFLFCNVNDVCFASRNDYS 42

RESULT 5

ADA20264

ID ADA20264 standard; protein; 79 AA.

XX

AC ADA20264;

XX

XX 20-NOV-2003 (first entry)

DT

XX

XX Human tumstatin deletion protein tum-5 amino acid sequence.

DE

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

KW metastasis; basement membrane organisation; type IV collagen network;

KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX

OS Homo sapiens.


```
XX PN WO2003059257-A2.
XX PD 24-JUL-2003.
XX PF 20-DEC-2002; 2002WO-US040938.
XX PR 21-DEC-2001; 2001US-00032221.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2003-587256/55.
XX DR N-PSDB; ADA20224.
XX PT New peptide, useful for preparing a composition for inhibiting tumor
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX PS Claim 94; SEQ ID NO 26; 240pp; English.
XX CC This invention relates to novel isolated proteins and their fragments
XX CC with anti-angiogenic properties. The invention also relates to the DNA
XX CC sequences which encode the novel proteins. A wide variety of diseases are
XX CC the result of undesirable angiogenesis. The formation of new capillaries
XX CC from pre-existing vessels is essential for tumor growth and metastasis.
XX CC Basement membrane organization is dependent on the assembly of a type IV
XX CC collagen network which may occur through the C-terminal globular non-
XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX CC forms are ubiquitously exhibited in human basement membranes. In the
XX CC present invention, cell surface receptors (in particular integrins) which
XX CC specifically bind anti-angiogenic proteins and peptides (in particular
XX CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX CC collagen) are disclosed. The proteins of the invention may inhibit tumor
XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in
XX CC endothelial cells and thus may exhibit cytostatic activity. The DNA
XX CC sequences of the invention may be useful in gene therapy. The present
XX CC invention is that of tum-5, an abridged form of the "tumstatin" protein of
XX CC the invention which was derived from the amino acid sequence of the alpha
XX CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does
XX CC not appear in the specification but was created by the indexer from
XX CC information given in the specification.
XX SQ Sequence 79 AA;
Query Match 84.1%; Score 122; DB 6; Length 79;
Best Local Similarity 92.3%; Pred. No. 5.9e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRFITMPFLFDNVNDVNFASRNDYS 27
DB 16 QRFITMPFLFCNVNDVNCVFASRNDYS 41
RESULT 6
AAU75608
ID AAU75608 standard; protein; 88 AA.
XX AC AAU75608;
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 82
/note= "Wild type Cys substituted with Ala"
WO200151523-A2.
19-JUL-2001.
08-JAN-2001; 2001WO-US000565.
07-JAN-2000; 2000US-00479118.
04-APR-2000; 2000US-00543371.
21-JUL-2000; 2000US-00625191.
(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
Kalluri R;
WPI; 2002-188037/24.
A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
treating disorders involving angiogenesis.
Claim 41; Page 153; 205pp; English.
The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
domain, having one or more of the characteristics selected from: (a) the
ability to bind alphavbeta3 integrin; (b) the ability to inhibit
proliferation of endothelial cells; and (c) the ability to cause
apoptosis of endothelial cells. Also described are the following: (1) use
of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
analogue or allelic variant in the preparation of a medicament for
treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
where the angiogenesis is mediated by one or more endothelial cell
integrins or one or more endothelial cell integrin subunits; or (b) by
promoting or inducing endothelial cell apoptosis in a tissue, where the
endothelial cell apoptosis is mediated by one or more endothelial cell
integrins or one or more endothelial cell integrin subunits; (2) use of
an antibody or peptide that specifically binds the alpha1, alpha2,
alpha3, alpha4, alpha6, alpha7, beta1 or beta2 subunit of integrin in the
preparation of a medicament for inhibiting angiogenesis or cell
proliferation; (3) use of an inhibitor, such as an antibody, antibody
fragment or peptide of receptor-mediated angiogenesis in the preparation
of a medicament for treating a proliferative disease in a vertebrate,
where the disease is characterised by angiogenesis that is mediated by
receptors to Arresten, Canstatin or Tumstatin and where the receptors
inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
the presence of a medicament for promoting angiogenesis in a tissue; and
(5) use of integrins in the preparation of a medicament for promoting or
inducing angiogenesis or cell proliferation in a tissue. The fragments
Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
or allelic variants are useful in the preparation of a medicament for
treating a disorder involving inhibiting angiogenesis in a tissue, where
the angiogenesis is mediated by one or more endothelial cell integrins or
one or more endothelial cell integrin subunits; or by promoting or
inducing endothelial cell apoptosis in a tissue, where the endothelial
cell apoptosis is mediated by one or more endothelial cell integrins or
one or more endothelial cell integrin subunits. The medicament is useful
in inhibiting tumour growth and for the regression of an established
tumour. The present sequence represents the amino acid sequence of human
type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
consists of residues 5-126 of Tumstatin
Sequence 88 AA;
Query Match 84.1%; Score 122; DB 5; Length 88;
Best Local Similarity 92.3%; Pred. No. 6.7e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRFITMPFLFDNVNDVNFASRNDYS 27
DB 26 QRFITMPFLFCNVNDVNCVFASRNDYS 51
```


CC ID33) does not appear in the specification but was created by the indexer
 CC from information given in the specification.

XX SQ Sequence 88 AA;

Query Match 84.1%; Score 122; DB 6; Length 88;

Best Local Similarity 92.3%; Pred. No. 6.7e-12; Mismatches 2; Indels 0; Gaps 0;

XX 2 QRTTTFPLFDNVNDVNFASRNDYS 27

Db 25 QRTTTFPLFCNVNDVNFASRNDYS 50

RESULT 9

ADA20272

ID ADA20272 standard; protein; 88 AA.

XX ADA20272;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX metastasis; basement membrane organisation; type IV collagen network;

XX C-terminal globular non-collagenous domain; NC1; type IV collagen;

XX cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX cytosolic; gene therapy; alpha 3 chain; tumstatin; human;

XX tumstatin 5-125-C-A; mutant; mutein.

XX Synthetic.

XX Homo sapiens.

XX Key

Location/Qualifiers

FT Misc-difference 81

FT /note= "Wild-type Cys substituted by Ala at position 125
 of full-length tumstatin"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor
 growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 34; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
 with anti-angiogenic properties. The invention also relates to the DNA
 sequences which encode the novel proteins. A wide variety of diseases are
 the result of undesirable angiogenesis. The formation of new capillaries
 from pre-existing vessels is essential for tumor growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 collagen network which may occur through the C-terminal globular non-
 collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 forms are ubiquitously exhibited in human basement membranes. In the
 present invention, cell surface receptors (in particular integrins) which
 specifically bind anti-angiogenic proteins and peptides (in particular
 the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 collagen) are disclosed. The proteins of the invention may inhibit tumour
 growth, angiogenic activity in mammalian tissue or protein synthesis in
 endothelial cells and thus may exhibit cytostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present
 sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
 the "tumstatin" protein of the invention which was derived from the amino
 acid sequence of the alpha 3 chain of human type IV collagen. Note: This
 sequence (Seq ID33) does not appear in the specification but was created
 CC by the indexer from information given in the specification.

XX SQ Sequence 88 AA;

Query Match 84.1%; Score 122; DB 6; Length 88;

Best Local Similarity 92.3%; Pred. No. 6.7e-12;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX 2 QRTTTFPLFDNVNDVNFASRNDYS 27

Db 25 QRTTTFPLFCNVNDVNFASRNDYS 50

RESULT 10

AAU75594

ID AAU75594 standard; protein; 124 AA.

XX AC AAU75594;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin 333.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;

XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US0000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00825191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-189037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 treating disorders involving angiogenesis.
 XX Example 33; Page; 205pp; English.
 XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 domain, having one or more of the characteristics selected from: (a) the
 ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 proliferation of endothelial cells; and (c) the ability to cause
 apoptosis of endothelial cells. Also described are the following: (1) use
 of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 where the angiogenesis is mediated by one or more endothelial cell
 integrins or one or more endothelial cell integrin subunits; or (b) by
 promoting or inducing endothelial cell apoptosis in a tissue, where the
 endothelial cell apoptosis is mediated by one or more endothelial cell
 integrins or one or more endothelial cell integrin subunits; (2) use of
 an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 preparation of a medicament for inhibiting angiogenesis or cell
 proliferation; (3) use of an inhibitor, such as an antibody, antibody

CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of
CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in
CC the specification but is derived from the wild type human Tumstatin
CC sequence given in figure 18A (see AAU75589)

XX
SQ Sequence 124 AA;

Query Match 84.1%; Score 122; DB 5; Length 124;
Best Local Similarity 92.3%; Pred. No. 1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDNVNDNFASRNDYS 27
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 69 QRFTTTPFLFCNVNDVCFASRNDYS 94

RESULT 11

ADA20258
ID ADA20258 standard; protein; 124 AA.

XX
AC ADA20258;

XX
DT 20-NOV-2003 (first entry)

XX
DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytotostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

XX
OS Homo sapiens.

XX
PN WO2003059257-A2.

XX
PD 24-JUL-2003.

XX
PF 20-DEC-2002; 2002WO-US040938.

XX
PR 21-DEC-2001; 2001US-00032221.

XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX
PI Kalluri R;

XX
XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

XX
XX New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX
PS Claim 94; SEQ ID NO 20; 240pp; English.

XX

CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
CC ID20) does not appear in the specification but was created by the indexer
CC from information given in the specification.

XX
SQ Sequence 124 AA;

Query Match 84.1%; Score 122; DB 6; Length 124;
Best Local Similarity 92.3%; Pred. No. 1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDNVNDNFASRNDYS 27
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 69 QRFTTTPFLFCNVNDVCFASRNDYS 94

RESULT 12

AAU75597

ID AAU75597 standard; protein; 132 AA.

XX
AC AAU75597;

XX
DT 08-MAY-2002 (first entry)

XX
DE Human type IV collagen alpha 3 chain mutant, Tum-2.

XX Human; type IV collagen alpha 3 chain; cytotostatic; antiangiogenic;
XX non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX
OS Homo sapiens.

XX
PN WO200151523-A2.

XX
PD 19-JUL-2001.

XX
PF 08-JAN-2001; 2001WO-US000565.

XX
PR 07-JAN-2000; 2000US-00479118.

XX
PR 04-APR-2000; 2000US-00543371.

XX
PR 21-JUL-2000; 2000US-00625191.

XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX
PI Kalluri R;

XX
XX WPI; 2002-188037/24.

XX
XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
PT treating disorders involving angiogenesis.

XX
PS Claim 31; Page 152; 205pp; English.

XX
CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI
CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphabeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 19A (see AAU75595)

XX SQ Sequence 132 AA;

Query Match 84.1%; Score 122; DB 5; Length 132;
 Best Local Similarity 92.3%; Pred. No. 1.1e-11;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFDNVNDVNFASRNDYS 27
 |||||
 Db 70 QRFTTMPFLFCNVNDVNCVFASRNDYS 95
 |||||

RESULT 13

ADA20261
 ID ADA20261 standard; protein; 132 AA.

XX AC

XX AC

XX AC

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCL1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytoskeletal; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX OS Homo sapiens.

XX OS

XX FN WO200305257-A2.

XX XX

XX PD 24-JUL-2003.

XX XX

XX PF 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor

XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 23; 240bp; English.

XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NCL) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCL domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.

XX SQ Sequence 132 AA;

Query Match 84.1%; Score 122; DB 6; Length 132;
 Best Local Similarity 92.3%; Pred. No. 1.1e-11;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTMPFLFDNVNDVNFASRNDYS 27
 |||||
 Db 69 QRFTTMPFLFCNVNDVNCVFASRNDYS 94
 |||||

RESULT 14

AAU75596

ID AAU75596 standard; protein; 191 AA.

XX AC

XX AC

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX XX

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3 (IV)NCL domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten;
 KW Tumstatin; angiogenesis; tumour; mutant.

XX OS Homo sapiens.

XX OS

XX PN WO200151523-A2.

XX XX

XX PD 19-JUL-2001.

XX XX

XX PF 08-JAN-2001; 2001WO-US0000565.

XX XX

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2002-188037/24.
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
PT treating disorders involving angiogenesis.
XX Example 32; Page; 205pp; English.
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of
CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in
CC the specification but is derived from the wild type human Tumstatin
CC sequence given in figure 18A (see AAU75589).
XX
SQ Sequence 191 AA;
Query Match 84.1%; Score 122; DB 5; Length 191;
Best Local Similarity 92.3%; Pred. No. 1.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRFTTTPFLFDNVNDVNFASRNDYS 27
DB 17 QRFTTTPFLFDNVNDVNCNFASRNDYS 42
RESULT 15
ADA20260
ID ADA20260 standard; protein; 191 AA.
XX
AC ADA20260;
XX
DT 20-NOV-2003 (first entry)
XX

DE Human tumstatin deletion protein tum-1 amino acid sequence.
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;
KW tumstatin N53.
XX OS Homo sapiens.
XX WO2003059257-A2.
XX 24-JUL-2003.
XX 20-DEC-2002; 2002WO-US040938.
XX 21-DEC-2001; 2001US-00032221.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2003-587256/55.
XX N-PSDB; ADA20224.
XX New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX Claim 94; SEQ ID NO 22; 240pp; English.
XX This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC "tumstatin" protein of the invention which was derived from the amino
CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This
CC sequence (Seq ID22) does not appear in the specification but was created
CC by the indexer from information given in the specification.
XX
SQ Sequence 191 AA;
Query Match 84.1%; Score 122; DB 6; Length 191;
Best Local Similarity 92.3%; Pred. No. 1.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRFTTTPFLFDNVNDVNFASRNDYS 27
DB 16 QRFTTTPFLFDNVNDVNCNFASRNDYS 41
Search completed: April 5, 2004, 06:58:32
Job time : 24.3196 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 16.9322 Seconds
(without alignments)
418.737 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTMBPFDVNDVDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 145 | 100.0 | 27 | 14 | US-10-032-221B-42 |
| 2 | 133 | 91.7 | 27 | 14 | US-10-032-221B-40 |
| 3 | 127 | 87.6 | 27 | 14 | US-10-032-221B-39 |
| 4 | 122 | 84.1 | 79 | 14 | US-10-032-221B-26 |
| 5 | 122 | 84.1 | 88 | 14 | US-10-032-221B-33 |
| 6 | 122 | 84.1 | 88 | 14 | US-10-032-221B-34 |
| 7 | 122 | 84.1 | 124 | 14 | US-10-032-221B-20 |
| 8 | 122 | 84.1 | 132 | 14 | US-10-032-221B-23 |
| 9 | 122 | 84.1 | 131 | 14 | US-10-032-221B-22 |
| 10 | 122 | 84.1 | 211 | 14 | US-10-270-877-46 |
| 11 | 122 | 84.1 | 211 | 14 | US-10-270-837-46 |
| 12 | 122 | 84.1 | 232 | 14 | US-10-206-699-304 |
| 13 | 122 | 84.1 | 244 | 14 | US-10-032-221B-30 |
| 14 | 106 | 73.1 | 229 | 14 | US-10-206-699-306 |
| 15 | 106 | 73.1 | 309 | 9 | US-09-925-297-496 |

| | | | | | | |
|----|-----|------|------|----|---------------------|--------------------|
| 16 | 105 | 72.4 | 229 | 14 | US-10-206-699-302 | Sequence 302, App |
| 17 | 105 | 72.4 | 229 | 14 | US-10-032-221B-2 | Sequence 2, Appli |
| 18 | 105 | 72.4 | 406 | 9 | US-09-925-302-507 | Sequence 507, App |
| 19 | 105 | 72.4 | 1669 | 15 | US-10-372-683-8 | Sequence 8, Appli |
| 20 | 101 | 69.7 | 22 | 14 | US-10-206-699-266 | Sequence 266, App |
| 21 | 101 | 69.7 | 25 | 14 | US-10-032-221B-37 | Sequence 37, Appli |
| 22 | 93 | 64.1 | 25 | 14 | US-10-032-221B-38 | Sequence 38, Appli |
| 23 | 91 | 62.8 | 22 | 14 | US-10-206-699-265 | Sequence 265, App |
| 24 | 91 | 62.8 | 1759 | 15 | US-10-369-493-7032 | Sequence 7032, App |
| 25 | 89 | 61.4 | 22 | 14 | US-10-206-699-267 | Sequence 267, App |
| 26 | 86 | 59.3 | 20 | 14 | US-10-206-699-289 | Sequence 289, App |
| 27 | 86 | 59.3 | 20 | 14 | US-10-032-221B-29 | Sequence 29, Appli |
| 28 | 83 | 57.2 | 46 | 9 | US-09-864-761-48095 | Sequence 48095, A |
| 29 | 83 | 57.2 | 1744 | 15 | US-10-369-493-5832 | Sequence 5832, Ap |
| 30 | 81 | 55.9 | 142 | 9 | US-09-864-761-38021 | Sequence 38021, A |
| 31 | 81 | 55.9 | 228 | 14 | US-10-206-699-307 | Sequence 307, App |
| 32 | 80 | 55.2 | 18 | 14 | US-10-206-699-254 | Sequence 254, App |
| 33 | 80 | 55.2 | 18 | 14 | US-10-206-699-260 | Sequence 260, App |
| 34 | 78 | 53.8 | 227 | 14 | US-10-206-699-303 | Sequence 303, App |
| 35 | 78 | 53.8 | 227 | 14 | US-10-032-221B-6 | Sequence 6, Appli |
| 36 | 78 | 53.8 | 430 | 9 | US-09-925-302-518 | Sequence 518, App |
| 37 | 78 | 53.8 | 459 | 15 | US-10-331-496A-27 | Sequence 27, Appli |
| 38 | 78 | 53.8 | 459 | 15 | US-10-372-683-30 | Sequence 30, Appli |
| 39 | 78 | 53.8 | 1712 | 10 | US-09-961-403-9 | Sequence 9, Appli |
| 40 | 74 | 51.0 | 18 | 14 | US-10-206-699-259 | Sequence 259, App |
| 41 | 74 | 51.0 | 22 | 14 | US-10-206-699-270 | Sequence 270, App |
| 42 | 72 | 49.7 | 18 | 14 | US-10-206-699-261 | Sequence 261, App |
| 43 | 71 | 49.0 | 22 | 14 | US-10-206-699-268 | Sequence 268, App |
| 44 | 70 | 48.3 | 18 | 14 | US-10-206-699-253 | Sequence 253, App |
| 45 | 70 | 48.3 | 20 | 14 | US-10-206-699-290 | Sequence 290, App |

ALIGNMENTS

RESULT 1

US-10-032-221B-42
; Sequence 42, Application US/10032221B
; Publication No. US2003014481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raguram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/136,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: P2 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length fms
; OTHER INFORMATION: tatin molecule, and aspartic acid has been substituted for the cy
; OTHER INFORMATION: steine residues at positions 79 and 85)
US-10-032-221B-42

Query Match 100.0%; Score 145; DB 14; Length 27;

```

; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T8 (amino acids 68-94 of SEQ ID NO:10; lysine has been s
; OTHER INFORMATION: d for the leucine residue at position 68 of the full-len
; OTHER INFORMATION: atin molecule)
; US-10-032-221B-39

Query Match      87.6%; Score 127; DB 14; Length 27;
Best Local Similarity 92.6%; Pred. No. 9e-13;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 KQRTTMPFLFDNVNDVNFASRNDYS 27
        |||||
DB      1 KQRTTMPFLFCNVNDVCNFSASRNDYS 27
        |||||

RESULT 4
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF US
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-015)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
; US-10-032-221B-26

Query Match      84.1%; Score 122; DB 14; Length 79;
Best Local Similarity 92.3%; Pred. No. 1.9e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTTMPFLFDNVNDVNFASRNDYS 27
        |||||
DB      16 QRFTTMPFLFCNVNDVCNFSASRNDYS 41
        |||||

RESULT 5
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1

```



```

US-10-032-221B-34
Query Match      84.1%; Score 122; DB 14; Length 88;
Best Local Similarity 92.3%; Pred.No. 2.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY    2 QRETTMPFLFDVNDVNDFASRNDYS 27
      |||||
DB    25 QRETTMPFLFCNVNDFCNFASRNDYS 50
      |||||

RESULT 7
US-10-032-221B-20
; Sequence 20, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match      84.1%; Score 122; DB 14; Length 124;
Best Local Similarity 92.3%; Pred.No. 3.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY    2 QRETTMPFLFDVNDVNDFASRNDYS 27
      |||||
DB    69 QRETTMPFLFCNVNDFCNFASRNDYS 94
      |||||

RESULT 8
US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224

```

```

; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 60/121,483
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match      84.1%; Score 122; DB 14; Length 132;
Best Local Similarity 92.3%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
Db      69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 9
US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match      84.1%; Score 122; DB 14; Length 191;
Best Local Similarity 92.3%; Pred. No. 5.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
Db      16 QRFTHMPFLFCNVNDVCFASRNDYS 41

RESULT 10
US-10-032-221B-21
; Sequence 46, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27

```

```

; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match      84.1%; Score 122; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
Db      69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 11
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match      84.1%; Score 122; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
Db      69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 12
US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27

```

; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 304
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: alpha 3 chain
US-10-206-699-304

Query Match 84.1%; Score 122; DB 14; Length 232;
Best Local Similarity 92.3%; Pred. No. 6.4e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27
|||||
Db 57 QRTTTPFLFCVNDVNCNFASRNDYS 82
|||||

RESULT 13

US-10-032-221B-10
; Sequence 10, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-10

Query Match 84.1%; Score 122; DB 14; Length 244;
Best Local Similarity 92.3%; Pred. No. 6.8e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27
|||||
Db 69 QRTTTPFLFCVNDVNCNFASRNDYS 94
|||||

RESULT 14

US-10-206-699-306
; Sequence 306, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 306
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: alpha 5 chain
US-10-206-699-306

Query Match 73.1%; Score 106; DB 14; Length 229;
Best Local Similarity 73.1%; Pred. No. 2e-08; 2; Indels 0; Gaps 0;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27
|||||
Db 55 RRTTTPFLFCVNDVNCNFASRNDYS 80
|||||

RESULT 15

US-09-925-297-496
; Sequence 496, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 496
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (247)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-496

Query Match 73.1%; Score 106; DB 9; Length 309;
Best Local Similarity 73.1%; Pred. No. 2.8e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27
|||||
Db 135 RRTTTPFLFCVNDVNCNFASRNDYS 160
|||||

Search completed: April 5, 2004, 07:36:07
Job time : 16.9322 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 6.3414 Seconds
(without alignments)
219.810 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 QRFTTTPFLFDVNDVDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pap.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pap.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pap.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pap.*
- 5: /cgn2_6/ptodata/2/iaa/6CTUS COMB.pap.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 122 | 84.1 | 211 | 4 | US-09-512-563C-46 |
| 2 | 122 | 84.1 | 218 | 2 | US-08-399-889-25 |
| 3 | 122 | 84.1 | 218 | 3 | US-09-167-364-25 |
| 4 | 122 | 84.1 | 218 | 3 | US-08-439-897-4 |
| 5 | 122 | 84.1 | 268 | 4 | US-09-589-927-6 |
| 6 | 122 | 84.1 | 268 | 4 | US-09-277-665-6 |
| 7 | 122 | 84.1 | 268 | 4 | US-09-589-987-6 |
| 8 | 121 | 83.4 | 471 | 2 | US-08-399-889-24 |
| 9 | 121 | 83.4 | 471 | 3 | US-09-167-364-24 |
| 10 | 121 | 83.4 | 471 | 3 | US-08-439-897-2 |
| 11 | 106 | 73.1 | 264 | 4 | US-09-589-927-10 |
| 12 | 106 | 73.1 | 264 | 4 | US-09-277-665-10 |
| 13 | 106 | 73.1 | 264 | 4 | US-09-589-987-10 |
| 14 | 105 | 72.4 | 260 | 4 | US-09-589-927-2 |
| 15 | 105 | 72.4 | 260 | 4 | US-09-277-665-2 |
| 16 | 105 | 72.4 | 260 | 4 | US-09-589-987-2 |
| 17 | 81 | 55.9 | 260 | 4 | US-08-589-927-12 |
| 18 | 81 | 55.9 | 260 | 4 | US-09-277-665-12 |
| 19 | 81 | 55.9 | 260 | 4 | US-09-589-987-12 |
| 20 | 78 | 53.8 | 258 | 4 | US-09-589-927-4 |
| 21 | 78 | 53.8 | 258 | 4 | US-09-277-665-4 |
| 22 | 78 | 53.8 | 258 | 4 | US-09-589-987-4 |
| 23 | 64 | 44.1 | 260 | 4 | US-09-589-927-8 |
| 24 | 64 | 44.1 | 260 | 4 | US-08-277-665-8 |
| 25 | 64 | 44.1 | 260 | 4 | US-09-589-987-8 |
| 26 | 55 | 37.9 | 411 | 4 | US-09-252-991A-23375 |
| 27 | 50 | 34.5 | 334 | 4 | US-09-252-991A-29546 |

28 46 31.7 1464 3 US-08-891-640-2 Sequence 2, Appli
29 46 31.7 1674 2 US-08-988-542C-12 Sequence 12, Appl
30 46 31.7 1674 4 US-09-554-467A-12 Sequence 12, Appl
31 45.5 31.4 165 4 US-09-540-236-2442 Sequence 2442, Ap
32 45.5 31.4 288 2 US-08-424-641B-11 Sequence 11, Appl
33 45.5 31.4 288 2 US-08-820-980-11 Sequence 11, Appl
34 45.5 31.4 288 2 US-08-836-439-11 Sequence 11, Appl
35 45 31.0 2548 4 US-09-172-422-1 Sequence 1, Appli
36 44 30.3 90 3 US-08-338-907-129 Sequence 129, App
37 44 30.3 90 4 US-09-218-207-129 Sequence 129, App
38 44 30.3 135 4 US-09-328-352-5695 Sequence 5695, Ap
39 44 30.3 165 4 US-09-134-000C-3704 Sequence 3704, Ap
40 44 30.3 205 4 US-09-489-039A-13421 Sequence 13421, A
41 44 30.3 334 4 US-09-543-681A-8177 Sequence 8177, Ap
42 44 30.3 1234 4 US-09-489-039A-8741 Sequence 8741, Ap
43 44 30.3 1694 1 US-08-494-168-2 Sequence 2, Appli
44 43.5 30.0 491 4 US-09-362-899-3 Sequence 3, Appli
45 43 29.7 309 4 US-09-107-532A-4994 Sequence 4994, Ap

ALIGNMENTS

RESULT 1
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512.563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 84.1%; Score 122; DB 4; Length 211;
Best Local Similarity 92.3%; Pred. No. 1.2e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFDVNDVDFASRNDYS 27
DB 69 QRFTTTPFLFCVNDVDFASRNDYS 94

RESULT 2
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT

; ORGANISM: Human
US-08-399-889-25

Query Match 84.1%; Score 122; DB 2; Length 218;
Best Local Similarity 92.3%; Pred. No. 1.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
|||||
Db 43 QRTTMPFLFCNVNDVCNFSRNDYS 68
|||||

RESULT 3

US-09-167-364-25

; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match 84.1%; Score 122; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. No. 1.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
|||||
Db 43 QRTTMPFLFCNVNDVCNFSRNDYS 68
|||||

RESULT 4

US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudec, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match 84.1%; Score 122; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. No. 1.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
|||||
Db 43 QRTTMPFLFCNVNDVCNFSRNDYS 68
|||||

RESULT 5

US-09-589-927-6
; Sequence 6, Application US/09589927

; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match 84.1%; Score 122; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 1.6e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
|||||
Db 93 QRTTMPFLFCNVNDVCNFSRNDYS 118
|||||

RESULT 6

US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match 84.1%; Score 122; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 1.6e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
|||||
Db 93 QRTTMPFLFCNVNDVCNFSRNDYS 118
|||||

RESULT 7

US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match 84.1%; Score 122; DB 4; Length 268;

Best Local Similarity 92.3%; Pred. No. 1.6e-11; Mismatches 0; Gaps 0; Indels 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27
Db 93 QRFTHMPFLFCNVDVDFASNDYS 118

RESULT 8

US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120

GENERAL INFORMATION:

; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match 83.4%; Score 121; DB 2; Length 471;
Best Local Similarity 88.5%; Pred. No. 4.6e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy .2 QRFTHMPFLFDVNDVDFASNDYS 27
Db 296 QRFTHMPFLFCNVDVDFASNDYS 321

RESULT 9

US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980

GENERAL INFORMATION:

; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match 83.4%; Score 121; DB 3; Length 471;
Best Local Similarity 88.5%; Pred. No. 4.6e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27
Db 296 QRFTHMPFLFCNVDVDFASNDYS 321

RESULT 10

US-09-439-897-2
; Sequence 2, Application US/09439897

; Patent No. 6277558

GENERAL INFORMATION:

; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 83.4%; Score 121; DB 3; Length 471;
Best Local Similarity 88.5%; Pred. No. 4.6e-11;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27
Db 296 QRFTHMPFLFCNVDVDFASNDYS 321

RESULT 11

US-09-589-927-10
; Sequence 10, Application US/09589927
; Patent No. 6432706

GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-10

Query Match 73.1%; Score 106; DB 4; Length 264;
Best Local Similarity 73.1%; Pred. No. 5.4e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27
Db 90 RRFTHMPFLFCNVDVDFASNDYS 115

RESULT 12

US-09-277-665-10
; Sequence 10, Application US/09277665
; Patent No. 6440729

GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-10

Query Match 73.1%; Score 106; DB 4; Length 264;
Best Local Similarity 73.1%; Pred. No. 5.4e-09;

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVDFASRNDYS 27
:|||||:|:|:|
Db 90 RRFSTMPFMCNINNVCFASRNDYS 115

RESULT 13

US-09-589-987-10
; Sequence 10, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-10

Query Match 73.1%; Score 106; DB 4; Length 264;
Best Local Similarity 73.1%; Pred. No. 5.4e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVDFASRNDYS 27
:|||||:|:|:|
Db 90 RRFSTMPFMCNINNVCFASRNDYS 115

RESULT 14

US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 72.4%; Score 105; DB 4; Length 260;
Best Local Similarity 73.1%; Pred. No. 7.6e-09;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVDFASRNDYS 27
:|||||:|:|:|
Db 86 RRFSTMPFMCNINNVCFASRNDYS 111

RESULT 15

US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665

CURRENT FILING DATE: 1999-03-26

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 260

TYPE: PRT

ORGANISM: Human

US-09-277-665-2

Query Match 72.4%; Score 105; DB 4; Length 260;

Best Local Similarity 73.1%; Pred. No. 7.6e-09;

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVDFASRNDYS 27

:|||||:|:|:|

Db 86 RRFSTMPFMCNINNVCFASRNDYS 111

Search completed: April 5, 2004, 07:07:26

Job time : 6.3414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:31:44 ; Search time 21 seconds
(without alignments)

1117.654 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRGDSGSPATWTRGF.....KAGELEKIISRCQVCMKRRH 244

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database :

PIR 78:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 244 | 100.0 | 1670 | 1 CGHU3B | collagen alpha 3(I) |
| 2 | 112 | 45.9 | 220 | 2 B49736 | collagen alpha 3(I) |
| 3 | 61 | 25.0 | 81 | 2 A49736 | collagen alpha 3(I) |
| 4 | 39 | 16.0 | 246 | 2 I48302 | collagen alpha 3(I) |
| 5 | 39 | 16.0 | 471 | 2 A39024 | collagen alpha 3(I) |
| 6 | 38 | 15.6 | 161 | 2 S49488 | collagen alpha 3(I) |
| 7 | 36 | 14.8 | 52 | 2 S69113 | collagen alpha 3(I) |
| 8 | 22 | 9.0 | 79 | 2 C43928 | probable collagen |
| 9 | 17 | 7.0 | 253 | 2 I48304 | collagen alpha 5(I) |
| 10 | 17 | 7.0 | 258 | 2 B61228 | collagen alpha 1(I) |
| 11 | 17 | 7.0 | 754 | 2 A55267 | collagen alpha 5(I) |
| 12 | 17 | 7.0 | 1669 | 1 CGHU4B | collagen alpha 1(I) |
| 13 | 17 | 7.0 | 1689 | 1 C3M54B | collagen alpha 1(I) |
| 14 | 17 | 7.0 | 1681 | 1 S22317 | collagen alpha 5(I) |
| 15 | 12 | 4.9 | 1752 | 2 A45407 | collagen alpha 3(I) |
| 16 | 11 | 4.5 | 58 | 2 B43928 | probable collagen |
| 17 | 11 | 4.5 | 1744 | 2 S40991 | collagen alpha 1(I) |
| 18 | 11 | 4.5 | 1761 | 2 T13990 | collagen type IV a |
| 19 | 10 | 4.1 | 1712 | 1 CGHU2B | collagen alpha 2(I) |
| 20 | 10 | 4.1 | 1747 | 2 A54121 | collagen alpha-4 c |
| 21 | 9 | 3.7 | 312 | 2 I48303 | collagen alpha 4(I) |
| 22 | 9 | 3.7 | 775 | 2 A61228 | collagen alpha 2(I) |
| 23 | 9 | 3.7 | 1707 | 2 A33526 | collagen alpha 2(I) |
| 24 | 8 | 3.3 | 781 | 2 S77782 | probable 1-phospho |
| 25 | 8 | 3.3 | 281 | 2 C69709 | superoxide dismuta |
| 26 | 8 | 3.3 | 1691 | 1 CGHU6B | collagen alpha 6(I) |
| 27 | 8 | 3.3 | 1775 | 2 A31893 | collagen alpha 1(I) |
| 28 | 7 | 2.9 | 49 | 2 S50999 | superoxide dismuta |
| 29 | 7 | 2.9 | 122 | 2 C72639 | hypothetical prote |

| | | | | | |
|-----|---|-----|------|----------|--------------------|
| 30 | 7 | 2.9 | 124 | 2 B42526 | B3R protein - vacc |
| 31 | 7 | 2.9 | 125 | 2 T08598 | probable diol dehy |
| 32 | 7 | 2.9 | 168 | 2 G72679 | hypothetical prote |
| 33 | 7 | 2.9 | 183 | 2 T47251 | complex I protein |
| 34 | 7 | 2.9 | 183 | 2 A36621 | NADH2 dehydrogenas |
| 35 | 7 | 2.9 | 187 | 2 B90191 | conserved hypothet |
| 36 | 7 | 2.9 | 200 | 2 A87671 | cytochrome c oxida |
| 37 | 7 | 2.9 | 202 | 2 S51097 | superoxide dismuta |
| 38 | 7 | 2.9 | 202 | 2 JC4396 | superoxide dismuta |
| 39 | 7 | 2.9 | 211 | 2 E90174 | superoxide dismuta |
| 40 | 7 | 2.9 | 211 | 2 S34616 | superoxide dismuta |
| 41 | 7 | 2.9 | 216 | 2 E82020 | ABC transporter AT |
| 42 | 7 | 2.9 | 216 | 2 E81247 | cell division App- |
| 43 | 7 | 2.9 | 233 | 1 DSBYN | superoxide dismuta |
| 44 | 7 | 2.9 | 237 | 2 H84035 | hypothetical prote |
| 45 | 7 | 2.9 | 247 | 2 AB0520 | conserved hypothet |
| 46 | 7 | 2.9 | 255 | 2 JG0179 | superoxide dismuta |
| 47 | 7 | 2.9 | 257 | 2 H82575 | 3-deoxy-manno-octu |
| 48 | 7 | 2.9 | 259 | 1 WMS28 | complement factor |
| 49 | 7 | 2.9 | 261 | 2 A34476 | collagen alpha 2(I |
| 50 | 7 | 2.9 | 264 | 2 E69897 | hypothetical prote |
| 51 | 7 | 2.9 | 267 | 2 AD1561 | B. subtilis Ycar p |
| 52 | 7 | 2.9 | 267 | 2 AE1204 | B. subtilis Ycar p |
| 53 | 7 | 2.9 | 343 | 2 A86241 | hypothetical prote |
| 54 | 7 | 2.9 | 354 | 2 C83577 | hypothetical prote |
| 55 | 7 | 2.9 | 370 | 2 S52699 | hypothetical prote |
| 56 | 7 | 2.9 | 409 | 2 S57689 | hypothetical prote |
| 57 | 7 | 2.9 | 419 | 2 S41607 | atrolysin A (EC 3. |
| 58 | 7 | 2.9 | 438 | 2 H89960 | conserved hypothet |
| 59 | 7 | 2.9 | 441 | 2 T38239 | hypothetical prote |
| 60 | 7 | 2.9 | 450 | 2 E70590 | 3-phosphoshikimate |
| 61 | 7 | 2.9 | 451 | 2 T30603 | perlecan homolog 2 |
| 62 | 7 | 2.9 | 453 | 2 S18804 | collagen alpha 4(I |
| 63 | 7 | 2.9 | 478 | 2 D72344 | DNA polymerase III |
| 64 | 7 | 2.9 | 550 | 1 HMI52 | hemagglutinin prec |
| 65 | 7 | 2.9 | 550 | 1 HMI53 | hemagglutinin prec |
| 66 | 7 | 2.9 | 550 | 1 HMI77 | hemagglutinin prec |
| 67 | 7 | 2.9 | 550 | 1 HMI80 | hemagglutinin prec |
| 68 | 7 | 2.9 | 550 | 1 HMI33 | hemagglutinin prec |
| 69 | 7 | 2.9 | 550 | 1 HMI89 | hemagglutinin prec |
| 70 | 7 | 2.9 | 550 | 1 HMI21 | hemagglutinin prec |
| 71 | 7 | 2.9 | 550 | 1 HMI98 | hemagglutinin prec |
| 72 | 7 | 2.9 | 550 | 1 HMI15 | hemagglutinin prec |
| 73 | 7 | 2.9 | 550 | 1 HMI86 | hemagglutinin prec |
| 74 | 7 | 2.9 | 550 | 2 JQ1153 | hemagglutinin prec |
| 75 | 7 | 2.9 | 550 | 2 JQ1154 | hemagglutinin prec |
| 76 | 7 | 2.9 | 550 | 2 JQ1155 | hemagglutinin prec |
| 77 | 7 | 2.9 | 550 | 2 JQ1156 | hemagglutinin prec |
| 78 | 7 | 2.9 | 565 | 1 HMI51 | hemagglutinin prec |
| 79 | 7 | 2.9 | 565 | 1 HMI53 | hemagglutinin prec |
| 80 | 7 | 2.9 | 565 | 1 HMI54 | hemagglutinin prec |
| 81 | 7 | 2.9 | 565 | 1 HMI55 | hemagglutinin prec |
| 82 | 7 | 2.9 | 565 | 1 HMI56 | hemagglutinin prec |
| 83 | 7 | 2.9 | 565 | 1 HMI57 | hemagglutinin prec |
| 84 | 7 | 2.9 | 565 | 1 HMI58 | hemagglutinin prec |
| 85 | 7 | 2.9 | 565 | 1 HMI59 | hemagglutinin prec |
| 86 | 7 | 2.9 | 565 | 1 HMI5T | hemagglutinin prec |
| 87 | 7 | 2.9 | 565 | 1 HMI5E | hemagglutinin - in |
| 88 | 7 | 2.9 | 565 | 2 S33703 | hemagglutinin prec |
| 89 | 7 | 2.9 | 566 | 1 HMI5H | hemagglutinin prec |
| 90 | 7 | 2.9 | 566 | 1 HMI5HA | hemagglutinin prec |
| 91 | 7 | 2.9 | 566 | 1 HMI5HM | hemagglutinin prec |
| 92 | 7 | 2.9 | 566 | 1 HMI5V | hemagglutinin prec |
| 93 | 7 | 2.9 | 566 | 1 HMI5DU | hemagglutinin prec |
| 94 | 7 | 2.9 | 567 | 1 HMI5V | collagen alpha 4(I |
| 95 | 7 | 2.9 | 623 | 2 A45137 | exotoxin A precurs |
| 96 | 7 | 2.9 | 638 | 2 C83503 | exotoxin A precurs |
| 97 | 7 | 2.9 | 681 | 2 I38755 | transcription fact |
| 98 | 7 | 2.9 | 695 | 2 S49228 | sodium-dependent p |
| 99 | 7 | 2.9 | 715 | 2 AC0018 | probable membrane |
| 100 | 7 | 2.9 | 725 | 2 T09395 | envelope polypepte |
| 101 | 7 | 2.9 | 1225 | 2 T13123 | DNA replication pr |
| 102 | 7 | 2.9 | 1324 | 2 T13123 | |

103 121 2 T00333
 104 183 2 F8366
 105 190 1 CGHUB
 106 1758 2 T29350
 107 1759 2 T29351
 108 1763 2 S16366
 109 1966 2 T08991
 110 3660 1 S02041
 111 20 2 S39419
 112 27 2 T01664
 113 35 2 I39969
 114 36 2 PH1753
 115 44 2 AD1753
 116 49 2 AF0716
 117 56 2 S61509
 118 64 2 S75543
 119 65 2 S14712
 120 66 2 T06597
 121 72 2 AB3399
 122 76 2 D82844
 123 80 2 T45103
 124 86 2 AF2255
 125 90 2 AF1969
 126 92 2 E83334
 127 92 2 JH0716
 128 93 2 T01876
 129 96 1 CCMP55
 130 101 2 S30493
 131 103 2 B90251
 132 104 2 T44890
 133 105 2 E72599
 134 106 2 H86901
 135 107 2 T42275
 136 110 2 B71524
 137 110 2 T46071
 138 112 2 AF2540
 139 113 2 C31769
 140 113 2 AH1784
 141 115 2 B72848
 142 119 2 T49363
 143 119 2 S10914
 144 126 2 B95415
 145 127 2 A72388
 146 128 2 S49637
 147 134 2 A81062
 148 134 2 G81807
 149 136 2 C83908
 150 139 2 F70657
 151 140 2 C49829
 152 143 2 S58429
 153 151 1 GGICPH
 154 151 2 D72706
 155 152 2 A53051
 156 153 2 A96972
 157 154 2 E84093
 158 157 2 T49554
 159 159 2 T04297
 160 161 2 S04917
 161 161 2 RGHUJA
 162 166 1 RGHUJA
 163 166 2 S04934
 164 166 2 A45751
 165 167 2 C90888
 166 167 2 J01797
 167 167 2 F85729
 168 169 2 H70377
 169 169 2 C95919
 170 169 2 C95394
 171 170 2 F82908
 172 173 2 AD3343
 173 175 2 T45900
 174 176 2 F69179
 175 178 2 F84792

176 6 2.5 180 2 A98248
 177 6 2.5 182 2 T19126
 178 6 2.5 183 2 B90092
 179 6 2.5 185 2 D87274
 180 6 2.5 188 2 T25883
 181 6 2.5 189 1 IVBOIA
 182 6 2.5 190 2 E95420
 183 6 2.5 193 2 B82063
 184 6 2.5 194 2 G75328
 185 6 2.5 198 2 JEO228
 186 6 2.5 199 2 A89774
 187 6 2.5 201 2 G50888
 188 6 2.5 201 2 B97956
 189 6 2.5 203 1 S65033
 190 6 2.5 203 2 A42710
 191 6 2.5 206 2 A75508
 192 6 2.5 207 2 G84157
 193 6 2.5 209 2 S07725
 194 6 2.5 209 2 E87589
 195 6 2.5 211 2 D37471
 196 6 2.5 212 2 T49680
 197 6 2.5 213 2 H83682
 198 6 2.5 214 2 F72664
 199 6 2.5 215 2 AE2177
 200 6 2.5 217 1 C64411

ALIGNMENTS

RESULT 1

CGHUB
 collagen alpha 3(IV) chain precursor, long splice form - human
 N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
 C:Species: Homo sapiens (man)
 C:Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 22-Jun-1999
 C:Accession: A54763; A43928; A4043; A45971; A39785
 R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Resders, S.T.
 J. Biol. Chem. 269, 23013-23017, 1994
 A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpressor
 A:Reference number: A54763; MUID:94364994; PMID:8083201
 A:Accession: A54763
 A:Molecule type: mRNA
 A:Residues: 1-1670 <NAR>
 A:Cross-references: GB:X80031; NID:G577563; PID:G577564
 A:Experimental source: kidney
 R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
 J. Clin. Invest. 89, 592-601, 1992
 A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al
 A:Reference number: A43928; MUID:92147878; PMID:1737849
 A:Accession: A43928
 A:Molecule type: mRNA
 A:Residues: 1331-1524, 'I', 1526-1670 <TUR>
 A:Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896
 A:Experimental source: kidney
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 267, 13780-13784, 1992
 A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpastur
 ction.
 A:Reference number: A44043; MUID:93015826; PMID:1400291
 A:Accession: A44043
 A:Molecule type: DNA; mRNA
 A:Residues: 1386-1670 <QUI>
 A:Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896
 A:Note: sequence extracted from NCBI backbone (NCBIP:115597)
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 269, 17358, 1994
 A:Reference number: A44738; MUID:94274734; PMID:8006044
 A:Contents: annotation; erratum; correction to intronic sequence in A44043
 R:Bernal, D.; Quinones, S.; Saus, J.
 J. Biol. Chem. 268, 12090-12094, 1993
 A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A:Reference number: A45971; MUID:93280184; PMID:8505332

A;Accession: A45971
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1427-1444 <BER>
A;Note: sequence extracted from NCBI backbone (NCBIIP:133363); sequence incorrectly identified
R;Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reeder, S.T.
Am. J. Hum. Genet. 49, 545-554, 1991
A;Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
A;Reference number: A39786; MUID:91353570; PMID:1882840
A;Accession: A39786
A;Molecule type: mRNA
A;Residues: 1453-1593, 'A', 1595-1670 <MOR>
A;Cross-references: GB:S55790; NID:G234418; PIDN:AA19637.1; PID:G234419
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
C;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-2q37
A;Intons: 1385/1; 1418/1; 1488/1; 1585/3; 1643/2 #status incomplete
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-li
C;Function:
A;Description: minor structural component of extracellular basement membrane in kidney
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
F;29-42/Domain: amino-terminal nonhelical, NH1 <NH1>
F;43-1438/Region: interrupted helical
F;791-793/Region: cell attachment (R-G-D) motif
F;996-998/Region: cell attachment (R-G-D) motif
F;1154-1156/Region: cell attachment (R-G-D) motif
F;1306-1308/Region: cell attachment (R-G-D) motif
F;1345-1347/Region: cell attachment (R-G-D) motif
F;1432-1434/Region: cell attachment (R-G-D) motif
F;1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NC1>
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;31-33-39-41-125-422-476-582-722-809-1387/Disulfide bonds: interchain #status predi
P;23/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1460-1548-1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F;1505-1511-1616-1622/Disulfide bonds: #status predicted
F;1570-1662-1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 244; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 2,1e-239;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60
Db 1427 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 1486

QY 61 LGTLGSCLOPRTTTPFLFCNVNDVCFASNDYSYWLSTPALMNMNAPITG 120
Db 1487 LGTLGSCLOPRTTTPFLFCNVNDVCFASNDYSYWLSTPALMNMNAPITG 1546

QY 121 RCTVCEGPAIAVHSQTTDIPCPHGWSLWKGFSFIFTSAGSGTQALASPGSCLE 180
Db 1547 RCTVCEGPAIAVHSQTTDIPCPHGWSLWKGFSFIFTSAGSGTQALASPGSCLE 1606

QY 181 EFRASPLECHGRGTCNYNSVFWLASLNLPDMFRKPIPSVTKAGELEKIIISRQVCM 240
Db 1607 EFRASPLECHGRGTCNYNSVFWLASLNLPDMFRKPIPSVTKAGELEKIIISRQVCM 1666

QY 241 KXRH 244
Db 1667 KXRH 1670

RESULT 2
A49736
collagen alpha 3(IV) chain, medium splice form - human (fragment)
N;Contains: collagen alpha 3(IV) chain, splice form GP-V
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 12-Nov-1999 #text_change 17-Mar-2000
C;Accession: A49736; D49736; S69111
R;Feng, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene. I
A;Reference number: A49736; MUID:94124597; PMID:8294492
A;Accession: A49736
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 169-220 <PEN1>
A;Accession: A49736
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 22-220 <FEN2>
A;Cross-references: GDB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wie
Eur. J. Biochem. 229, 754-760, 1995
A;Title: Characterization and expression of multiple alternatively spliced transcripts c
utoc antigen and one of its alternative forms.
A;Reference number: S69111; MUID:95278230; PMID:7758473
A;Accession: S69111
A;Molecule type: mRNA
A;Residues: 1-45, 169-204, 'L', 206-220 <PEN>
C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.
C;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-2q37
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status predi
F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status
F;22-20/Domain: carboxyl-terminal nonhelical, NCI <NC1>
F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>
Query Match 45.9%; Score 112; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.4e-106;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60
Db 10 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 69

QY 61 LGTLGSCLOPRTTTPFLFCNVNDVCFASNDYSYWLSTPALMNMNAPITG 112
Db 70 LGTLGSCLOPRTTTPFLFCNVNDVCFASNDYSYWLSTPALMNMNAPITG 121

RESULT 3
A49736
collagen alpha 3(IV) chain, short splice form - human (fragment)
N;Contains: collagen alpha 3(IV) chain, splice form GP-III
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999
C;Accession: A49736; D49736; S69112
R;Feng, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene. I
A;Reference number: A49736; MUID:94124597; PMID:8294492
A;Accession: A49736
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 71-81 <FEN1>
A;Accession: A49736
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 22-81 <FEN2>

A;Cross-references: GB:U02520; NID:9408895; PIDN:AAAI8943.1.; PID:g408896
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank
R;Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A;Reference number: A45971; MUID:93280184; PMID:8505332
A;Accession: B45971
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 71-81 <BER>
A;Cross-references: PIDN:AA27014.1.; PID:g385563
A;Note: sequence extracted from NCBI backbone (NCBIP:133955); sequence incorrectly identified
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wied
Eur. J. Biochem. 229, 754-760, 1995
A;Title: Characterization and expression of multiple alternatively spliced transcripts of
autoantigen and one of its alternative forms.
A;Reference number: S69111; MUID:95278230; PMID:7758473
A;Accession: S69112
A;Molecule type: mRNA
A;Residues: 1-45,71-81 <PEN>
C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.
C;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-2q37
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
F;1-81/Product: collagen alpha 3(IV) chain, short splice form (fragment) #status predict
F;1-45,71-81/Product: collagen alpha 3 (IV) chain, splice form GP-III (fragment) #status
F;22-81/Domain: carboxyl-terminal nonhelical, NCI <NCI>

Query Match 25.0%; Score 61; DB 2; Length 81;
Best Local Similarity 100.0%; Pred. No. 2e-54;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKRGDSSPATWTRGTFVTRHSQTTAIPSCPGTVPVLSGFSFLVQGNORAHGD 60
|||||
Db 10 GLKRGDSSPATWTRGTFVTRHSQTTAIPSCPGTVPVLSGFSFLVQGNORAHGD 69
|||||

Qy 61 L 61
70 L 70

RESULT 4
148302
collagen alpha 3(IV) chain - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
C;Accession: 148302; S47278
R;Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994
A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
A;Reference number: A54979; MUID:95050957; PMID:7962065
A;Accession: 148302
A;Status: Preliminary
A;Molecule type: mRNA
A;Residues: 1-246 <RES>
A;Cross-references: EMBL:Z35166; NID:9535197; PID:g535198
C;Superfamily: collagen alpha 1(IV) chain

Query Match 16.0%; Score 39; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 ECHGRGTCNYSNSYSFWLASLNERFRKPIPTVKAG 227
|||||
Db 191 ECHGRGTCNYSNSYSFWLASLNERFRKPIPTVKAG 229
|||||

RESULT 5
A39024
collagen alpha 3(IV) chain - bovine (fragment)

C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
A;Accession: A39024; S20672; S17802; A3167; C39419; S13747; S20815
R;Morrison, K.E.; Germino, G.G.; Reeders, S.T.
J. Biol. Chem. 266, 34-39, 1991
A;Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
A;Reference number: A39024; MUID:91093146; PMID:1985905
A;Accession: A39024
A;Molecule type: mRNA
A;Residues: 1-471 <MOR>
A;Cross-references: EMBL:M63139; NID:g162886; PIDN:AA62708.1.; PID:g162887
R;Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J. Biol. Chem. 262, 7874-7877, 1987
A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane c
A;Reference number: S18432; MUID:87222419; PMID:2438283
A;Accession: S20672
A;Molecule type: protein
A;Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
R;Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J. Biol. Chem. 265, 5465-5469, 1990
A;Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type
A;Reference number: A35167; MUID:90202779; PMID:2318822
A;Accession: A35167
A;Molecule type: protein
A;Residues: 236-258 <GUN>
R;Gunwar, S.; Ballister, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
J. Biol. Chem. 266, 15318-15324, 1991
A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll
A;Reference number: A39419; MUID:91332055; PMID:1869555
A;Accession: C39419
A;Molecule type: protein
A;Residues: 236-255 <GU2>
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F;1-238/Domain: collagenous (fragment) #status predicted <COL>
F;233-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
F;233-353/Domain: repeat NCI #status predicted <NCI1>
F;354-471/Domain: repeat NCI #status predicted <NCI2>
F;232,238/Modified site: hydroxyproline (Pro) #status experimental
F;306-312,417-423/Disulfide bonds: #status predicted

Query Match 16.0%; Score 39; DB 2; Length 471;
Best Local Similarity 100.0%; Pred. No. 2.1e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 MAPITGRALEPIYISRCTVCEGPAIAIAVHSQTTDIPPCP 145
|||||
Db 334 MAPITGRALEPIYISRCTVCEGPAIAIAVHSQTTDIPPCP 372
|||||

RESULT 6
S49488

collagen alpha 3(IV) chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C;Accession: S49488
R;Oberbaumer, I.
submitted to the EMBL Data Library, October 1994
A;Description: Cloning of the NCI domains fo the minor collagen IV chains of mouse via F
ells.
A;Reference number: S49487
A;Accession: S49488
A;Status: Preliminary
A;Molecule type: mRNA
A;Residues: 1-161 <OBE>
A;Cross-references: EMBL:X82205; NID:g559472; PIDN:CAA57689.1; PID:g559916

C:Superfamily: collagen alpha 1(IV) chain

Query Match 15.6%; Score 38; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 8.9e-31;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYNSFWLASLNPMPKPIPTVKA 226
DB 124 ECHGRGTCNYNSYNSFWLASLNPMPKPIPTVKA 161

RESULT 7

S69113 collagen alpha 3(IV) chain, splice form GP-III/IV/V - human

C:Species: Homo sapiens (man)

C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 12-Nov-1999

C:Accession: S69113

R:Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wie

Eur. J. Biochem. 229, 754-760, 1995

A:Title: Characterization and expression of multiple alternatively spliced transcripts c

ucotagen and one of its alternative forms.

A:Reference number: S69111; MUID:95278230; PMID:7758473

A:Accession: S69113

A:Molecule type: mRNA

A:Residues: 1-52 <PEN>

C:Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.

C:Superfamily: collagen alpha 1(IV) chain

Query Match 14.8%; Score 36; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 3.7e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKGGSGPATWTRGTFVTRISQTTAIPSCPE 36
DB 10 GLKGKGGSGPATWTRGTFVTRISQTTAIPSCPE 45

RESULT 8

C43928

Probable collagen alpha 3(IV) chain - sheep (fragments)

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

C:Accession: C43928

R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al

A:Reference number: A43928; MUID:92147878; PMID:1737849

A:Accession: C43928

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-79 <TUR>

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 9.0%; Score 22; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 9e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYNSFWLASL 210
DB 27 ECHGRGTCNYNSYNSFWLASL 48

RESULT 9

I48304

collagen alpha 5(IV) chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999

C:Accession: I48304; S47280

R:Miner, J.H.; Sares, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48304

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-253 <RES>

A:Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202

C:Superfamily: collagen alpha 1(IV) chain

Query Match 7.0%; Score 17; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 94 VCNFASRNDYSYWLSTP 110

RESULT 10

B61228

collagen alpha 1(IV) chain - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999

C:Accession: B61228

R:Yamaguchi, N.; Sato, N.; Ko, J.S.; Niomiya, Y.

Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991

A:Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothe

A:Reference number: A61228; MUID:92010685; PMID:1717398

A:Accession: B61228

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-258 <YAM>

C:Superfamily: collagen alpha 1(IV) chain

Query Match 7.0%; Score 17; DB 2; Length 258;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 99 VCNFASRNDYSYWLSTP 115

RESULT 11

A55267

collagen alpha 5(IV) chain - dog (fragment)

C:Species: Canis lupus familiaris (dog)

C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999

C:Accession: A55267

R:Zheng, K.; Thorne, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994

A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-li

en type IV.

A:Reference number: A55267; MUID:94224868; PMID:8171024

A:Accession: A55267

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>

A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548

C:Superfamily: collagen alpha 1(IV) chain

Query Match 7.0%; Score 17; DB 2; Length 754;
Best Local Similarity 100.0%; Pred. No. 7.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 602 VCNFASRNDYSYWLSTP 618

RESULT 12

CGHU4B

collagen alpha 1(IV) chain precursor - human

N:Alternate names: procollagen alpha 1(IV) chain

C:Species: Homo sapiens (man)

C>Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999
C/Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
R/Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A/Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen
A/Reference number: S16876; MUID:89340433; PMID:2701944
A/Accession: S16876
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1669 <SO1>
A/Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988
R/Soininen, R.; Huotari, M.; Hosikaka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A/Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A/Reference number: A9690; MUID:89034231; PMID:3182844
A/Accession: A32117
A/Molecule type: DNA
A/Residues: 1-28 <SO12>
A/Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233
R/Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A/Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane
A/Reference number: S02738; MUID:89030632; PMID:2846280
A/Accession: S02738
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-6 'L', 8-28 <POE>
A/Cross-references: EMBL:X12784; NID:G30072
R/Brazeal, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A/Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
A/Reference number: S00048; MUID:88029471; PMID:3311751
A/Accession: S00048
A/Molecule type: mRNA
A/Residues: 1-318, A', 320-944 <BR1>
A/Cross-references: EMBL:X05561; NID:G30066; PIDN:CAR29075.1; PID:G30067
A/Accession: S25826
A/Molecule type: protein
A/Residues: 271-318, A', 320-554 <BR2>
R/Glanville, R.W.; Qian, R.Q.; Steibold, B.; Risteli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1995
A/Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (78
A/Reference number: A23115; MUID:86004708; PMID:4043082
A/Accession: A23115
A/Molecule type: protein
A/Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>
A/Experimental source: placenta
A/Note: The amino end of the mature form is blocked
R/Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
FEBS Lett. 225, 188-194, 1997
A/Title: Complete primary structure of the alpha(1)-chain of human basement membrane (ty
A/Reference number: S00207; MUID:88083584; PMID:3691802
A/Accession: S00207
A/Molecule type: mRNA
A/Residues: 244-530 <SO13>
A/Cross-references: EMBL:X00706; NID:G29548; PIDN:CAA68698.1; PID:G29549
R/Ebte, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A/Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
A/Reference number: S39614; MUID:94038963; PMID:8223498
A/Accession: S39614
A/Molecule type: protein
A/Residues: 371-554 <EBL>
R/Babel, W.; Glanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A/Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid se
A/Reference number: A02863; MUID:85003629; PMID:6434307
A/Accession: A02863
A/Molecule type: protein
A/Residues: 534-718, D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-
A/Experimental source: placenta
R/Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A/Title: Peptin fragments of human placental basement-membrane collagens showing interrui
A/Reference number: S16908; MUID:82005835; PMID:6792033
A/Accession: A58517
A/Molecule type: protein
A/Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-
R/MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A/Title: Isolation and characterization of pepsin-solubilized human basement membrane (t
A/Reference number: S16910; MUID:84053346; PMID:6416291
A/Accession: S16910
A/Molecule type: protein
A/Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-
A/Experimental source: placenta
R/Philajantiemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; P
J. Biol. Chem. 260, 7681-7687, 1985
A/Title: cDNA clones coding for the pro-alpha-1(IV) chain of human type IV procollagen r
A/Reference number: S01466; MUID:85207819; PMID:2581969
A/Accession: S01466
A/Molecule type: mRNA
A/Residues: 1256-1669 <PIH>
A/Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424
R/Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kafalides, N.A.;
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A/Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
A/Reference number: S16879; MUID:85216555; PMID:2582422
A/Accession: S16879
A/Molecule type: mRNA
A/Residues: 1259-1669 <BRI>
A/Cross-references: EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PID:G180818
R/Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
Eur. J. Biochem. 147, 217-224, 1985
A/Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A/Reference number: A02864; MUID:85127033; PMID:2578961
A/Accession: S19091
A/Molecule type: protein
A/Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',
R/Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A/Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
A/Reference number: S02550; MUID:89005112; PMID:2844531
A/Contents: annotation; disulfide bonds
C/Genetics:
A/Gene: GDB:COL4A1
A/Cross-references: GDB:119791; OMIM:120130
A/Map position: 13q34-13q34
A/Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/
1/; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
C/Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(
C/Complex: among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
r-trimer associations in the interrupted helical domain (with disulfide and desmosine cr
C/Function:
A/Description: structural component of extracellular basement membrane
C/Superfamily: collagen alpha 1(IV) chain
C/Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication
F.1-26/Domain: signal sequence #status predicted <SIG>
F.27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F.29-162/Domain: amino-terminal nonhelical, 7S <7SD>
F.163-1440/Domain: interrupted helical <COL>
F.141-452/Region: integrin binding #status experimental
F.597-599/Region: cell attachment (R-G-D) motif
F.917-919/Region: cell attachment (R-G-D) motif
F.968-970/Region: cell attachment (R-G-D) motif
F.1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F.1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR>
F.1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>
F.27/Modified site: blocked amino end (Aia) (in mature form) #status experimental
F.31.36.39, 41, 125.434, 467, 470/Disulfide bonds: interchain #status predicted
F.31.36.39, 41, 125.434, 467, 470/Disulfide bonds: interchain #status predicted
F.45.48, 78, 90, 129, 156, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 543, 573, 582, 61
1081, 1084, 1099, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 134
F.45.48, 78, 90, 129, 156, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 543, 573, 582, 61
99, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 1340, 1356, 1371,
F.54, 63, 75, 84, 87, 96, 102, 105, 108, 111, 117, 120, 123, 138, 141, 147, 150, 153, 159, 167, 178, 181, 184,

419, 422, 425, 439, 445, 448, 451, 479, 485, 491, 494, 503, 512, 518, 524, 530, 546, 549, 552, 555, 561, 567
 9, 745, 748, 751, 754, 763/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:126/Binding site: carbonylate (Asn) (covalent) #status experimental
 F:127/Modified site: allylsine (Lys) #status predicted
 F:172,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical
 F:272,645,839/Modified site: 4-hydroxyproline (Pro) #status atypical
 F:446-447/Cleavage site: Gly-Ile (Gelatinase B) #status predicted
 F:766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899,90
 23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,1256,
 431,1437/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:1120,1268/Binding site: carbonylate (Lys) (covalent) (partial) #status experimental
 F:1214,1424/Modified site: 3-hydroxyproline (Pro) #status absent
 F:1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 7.0%; Score 17; DB 1; Length 1669;
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
 1510 VCNFASRNDYSYWLSTP 1526

Db

RESULT 13
 CGMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000
 C:Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A31766; S19
 R:Nuthukumar, G.; Blumberg, B.; Kurkinen, M.
 J. Biol. Chem. 264, 6310-6317, 1989
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff
 A:Reference number: A33525; MUID:89197932; PMID:2703490
 A:Accession: A33525
 A:Molecule type: mRNA
 A:Residues: 1-1669 <EMBL>
 A:Cross-references: EMBL:J04694; NID:G556296; PIDN:AAA50292.1; PID:G556297
 R:Wood, L.; Theriault, N.; Vogeli, G.
 FEBS Lett. 227, 5-8, 1988
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al
 A:Reference number: S01454; MUID:88112221; PMID:339568
 A:Accession: S01454
 A:Molecule type: mRNA
 A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71
 A:Cross-references: EMBL:X06777
 R:Killen, P.D.; Burbeio, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain
 A:Reference number: A28066; MUID:88243724; PMID:3379041
 A:Accession: A28066
 A:Molecule type: mRNA
 A:Residues: 1-129 <K1>
 A:Cross-references: EMBL:J03758; NID:G192659; PIDN:AAA37439.1; PID:G192670
 R:Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A:Reference number: A02864; MUID:85127033; PMID:2578961
 A:Accession: A02864
 A:Molecule type: mRNA
 A:Residues: 1276-1659 <OBE>
 A:Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
 Gene 43, 301-304, 1986
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
 A:Reference number: A25636; MUID:86301886; PMID:3755692
 A:Accession: A25636
 A:Molecule type: mRNA
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
 A:Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287

A:Note: the authors translated the codon CAG for residue 1374 as Arg
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj
 J. Biol. Chem. 262, 8496-8499, 1987
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
 A:Reference number: A94680; MUID:87250460; PMID:3597383
 A:Accession: A29301
 A:Molecule type: mRNA
 A:Residues: 1441-1669 <KUR>
 A:Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115
 R:Killen, P.D.; Burbeio, P.D.; Martin, G.R.; Yamada, Y.
 J. Biol. Chem. 263, 12310-12314, 1988
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
 A:Reference number: S19079; MUID:88315019; PMID:2842328
 A:Accession: S19079
 A:Molecule type: DNA
 A:Residues: 1-28 <K12>
 A:Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G456503
 R:Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
 J. Biol. Chem. 263, 19274-19277, 1988
 A:Title: Head-to-head arrangement of murine type IV collagen genes.
 A:Reference number: A92702; MUID:89066738; PMID:3198626
 A:Accession: A32003
 A:Molecule type: DNA
 A:Residues: 1-28 <KAY>
 A:Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449
 R:Burbeio, P.D.; Martin, G.R.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
 A:Reference number: A94220; MUID:89071759; PMID:3200851
 A:Accession: A31766
 A:Molecule type: DNA
 A:Residues: 1-28 <BUR>
 A:Cross-references: EMBL:M02333; NID:G340878; PIDN:AAA51625.1; PID:G535668
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
 J. Biol. Chem. 261, 6654-6657, 1986
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
 A:Reference number: S19094; MUID:86196099; PMID:3009468
 A:Accession: S19094
 A:Molecule type: DNA
 A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>
 A:Cross-references: EMBL:M13027
 R:Schuppan, D.; Timpl, R.; Glanville, R.W.
 FEBS Lett. 115, 297-300, 1980
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
 A:Reference number: S16909; MUID:80246483; PMID:6772473
 A:Accession: S16909
 A:Molecule type: protein
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 123, 505-512, 1982
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin
 A:Reference number: A25991; MUID:82186723; PMID:6804236
 A:Accession: A25991
 A:Molecule type: protein
 A:Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 100
 51, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119,
 A:Accession: B25991
 A:Molecule type: protein
 A:Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1234,
 3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1329
 R:Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 139, 401-410, 1984
 A:Title: Subunit structure and assembly of the globular domain of basement-membrane coll
 A:Reference number: S17801; MUID:84132058; PMID:6698021
 A:Accession: S17801
 A:Molecule type: protein
 A:Residues: 1435-1443 <WEB>
 C:Genetics:
 A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
 A:Note: the list of introns may be incomplete
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular mat
 P:1-27/Domain: signal sequence #status predicted <SIG>

F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F:28-162/Domain: 78 <7SD>
F:163-1440/Domain: collagenous, triple helix <COL>
F:597-599/Region: cell attachment (R-G-D) motif
F:781-783/Region: cell attachment (R-G-D) motif
F:917-919/Region: cell attachment (R-G-D) motif
F:968-970/Region: cell attachment (R-G-D) motif
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1441-1522/Region: duplication
F:1553-1669/Region: duplication
F:31,36,39,41,43,46,47,470/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbonyldrate (Asn) (covalent) #status predicted
F:971,974,977,986,989,1001,1022,1031,1037,1040,1055,1060,1063,1075,1078,1090,1121,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxypro
F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental
F:1505-1511,1516-1622/Disulfide bonds: #status predicted

Query Match 7.0%; Score 17; DB 1; Length 1669;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
Db 1510 VCNFASRNDYSYWLSTP 1526
|||||

RESULT 14
S22917
N:Collagen alpha 5(IV) chain precursor, renal splice form - human
N:Alternate names: procollagen alpha 5(IV) chain
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence, revision 27-Feb-1997 #text change 21-Jul-2000
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
R:Zhou, J.; Hert, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi
n Alport syndrome patient.
A:Reference number: S22917; MUID:92316923; PMID:1352287
A:Accession: S22917
A:Molecule type: mRNA
A:Residues: 1-967 <ZH0>
A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234
R:Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A:Title: Structure of the human type IV collagen COL4A5 gene.
A:Reference number: A54365; MUID:94165049; PMID:8120014
A:Accession: A54365
A:Molecule type: DNA
A:Residues: 1-922 <ZH2>
A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAC27816.1; PID
R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurilla, P.; de Paeppe, A.; Tryggvas
Science 261, 1167-1169, 1993
A:Title: Deletion of the paired alphas(IV) and alphas(IV) collagen genes in inherited sm
A:Reference number: A57079; MUID:93361972; PMID:8356449
A:Accession: A57079
A:Molecule type: DNA
A:Residues: 1-27 <ZH4>
A:Cross-references: GB:Z37153; NID:G587203; PIDN:CAA85512.1; PID:G587204
R:Pihtlajaniemi, T.; Pohjola, E.R.; Myers, J.C.
J. Biol. Chem. 265, 13758-13766, 1990
A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin
A:Reference number: A37122; MUID:90337990; PMID:2380186
A:Accession: A37122
A:Molecule type: mRNA
A:Residues: 84-439, 'GS', 442-624, 'LALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
A:Cross-references: GB:J05558; EMBL:M58526; NID:G1314209
A:Note: Submitted to the EMBL Data Library, February 1991
A:Note: the authors translated the codon GCC for residue 115 as Val
R:Renieri, A.; Seri, M.; Myers, J.C.; Pihtlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
Hum. Mol. Genet. 1, 127-129, 1992
A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in

A:Reference number: I54317; MUID:93244772; PMID:1363780
A:Accession: I54317
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 313-324, 'E', 326-330 <REN>
A:Cross-references: GB:S59334; NID:G299946; PIDN:AAD13909.1; PID:G4261609
R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hooyhtyae, M.; Shows, T.B.; Tryggvason, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidney
A:Reference number: A34850; MUID:90160375; PMID:1689491
A:Accession: A34850
A:Molecule type: mRNA
A:Residues: 914-1264, 1271-1691 <HOS>
A:Cross-references: EMBL:M31115; NID:G180824; PIDN:AAA52045.1; PID:G180825
R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.
Genomics 9, 1-9, 1991
A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
A:Reference number: A37969; MUID:91169491; PMID:2004755
A:Accession: S18950
A:Molecule type: DNA
A:Residues: 924-1264, 1271-1691 <ZH3>
A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB
8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:G177922; PIDN:AAA51558.1; PID
R:Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M
Kidney Int. 44, 1316-1321, 1993
A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
A:Reference number: I56971; MUID:94133540; PMID:8301933
A:Accession: I56971
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1258-1276 <GUO1>
A:Cross-references: GB:G69168; NID:G545095; PIDN:AAC60612.1; PID:G545096
A:Note: kidney splice form
A:Accession: I76598
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1284-1291, 'FLGYIACLV', <GUO2>
A:Cross-references: GB:G69169; NID:G545097; PIDN:AAC60613.1; PID:G545098
A:Note: frameshift mutation in patient with Alport syndrome
R:Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; So
Am. J. Hum. Genet. 46, 1024-1033, 1990
A:Title: Molecular cloning of alphas(IV) collagen and assignment of the gene to the regi
A:Reference number: A35335; MUID:90252791; PMID:2339699
A:Accession: A35335
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1448-1477 <MYE>
R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Yos
Kidney Int. 46, 1307-1314, 1994
A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord
A:Reference number: I56975; MUID:95156893; PMID:7853788
A:Accession: I56975
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1595-1602 <NAK>
A:Cross-references: GB:S75903; NID:G913882; PIDN:AAB33374.1; PID:G913883
A:Note: premature termination mutation from a patient with Alport syndrome; one other mu
R:Leimink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;
Genomics 17, 485-489, 1993
A:Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo
A:Reference number: I54188; MUID:94010948; PMID:8406498
A:Accession: I54188
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1604-1607, 'VHDYKVC', <LEM>
A:Cross-references: GB:S65767; NID:G425563; PIDN:AAD13967.1; PID:G4261667
A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations ar
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A5; ATS
A:Cross-references: GDB:120596; OMIM:303630
A:Map position: Xq22-Xq22

F:129/Modified site: allylsine (Lys) #status predicted

Query Match 4.9%; Score 12; DB 2; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRDYSYWLST 99
| | | | | | | | | | | | | | | | | | | | | |
DB 1597 ASRDYSYWLST 1608

RESULT 16
B43928
C:Probable collagen alpha 5(IV) chain - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: B43928
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be
A:Reference number: A43928; MUID:92147878; PMID:1737849
A:Accession: B43928
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-58 <TUR>
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 4.5%; Score 11; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYY 199
| | | | | | | | | | | | | | | | | | | | | |
DB 6 ECHGRGTCNYY 16

RESULT 17
S40991
collagen alpha 1(IV) chain precursor - Caenorhabditis elegans
N:Alternate names: protein K04H4.1
C:Species: Caenorhabditis elegans
C:Date: 03-May-1994 #sequence_revision 02-Aug-1994 #text_change 13-Aug-1999
C:Accession: S40991; S4442; S13651; B34476
R:Ainscough, R.
submitted to the EMBL Data Library, October 1993
A:Reference number: S40991
A:Accession: S40991
A:Molecule type: DNA
A:Residues: 1-1744 <AIN>
A:Cross-references: EMBL:Z27078; NID:g414627; PID:g414628
R:Kramer, J.M.
submitted to the EMBL Data Library, December 1990
A:Reference number: S44442
A:Accession: S44442
A:Molecule type: DNA
A:Residues: 1-129, 'GPPGMPGLAGPGQSGQNGNRPGLSGPPGEGVNSQGRKGVKESGRSGVPGLP', 208
15, 'D', 817-1260, 'P', 1262-1707, 'P', 1709-1744 <KRA>
A:Cross-references: EMBL:X56979; NID:g6675; PIDN:CAA40299.1; PID:g6676
R:Guo, X.; Johnson, J.J.; Kramer, J.M.
Nature 349, 707-709, 1991
A:Title: Embryonic lethality caused by mutations in basement membrane collagen
A:Reference number: S13651; MUID:91141582; PMID:1996137
A:Accession: S13651
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-129, 'GPPGMPGLAGPGQSGQNGNRPGLSGPPGEGVNSQGRKGVKESGRSGVPGLP', 208
15, 'D', 817-1260, 'P', 1262-1515 <GUI>
A:Cross-references: EMBL:X56979
R:Guo, X.; Kramer, J.M.
J. Biol. Chem. 264, 17574-17582, 1989
A:Title: The two Caenorhabditis elegans basement membrane collagen (type IV) collagen genes
A:Reference number: A34476; MUID:90008929; PMID:2793871

A;Accession: B34476
A;Molecule type: DNA
A;Residues: 1432-1499, 'Q', 1501-1707, 'P', 1709-1744 <GU2>
A;Cross-references: EMBL:J05067; NID:G156255; PIDN:AAB59179.1; PID:G156256
C;Genetics:
A;Gene: c1b-2; emb-9
A;Map position: 3
A;Introns: 23/2; 79/1; 152/2; 288/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F;43-1515/Domain: collagenous, triple helix #status predicted <COL>
F;93-95/Region: cell attachment (R-G-D) motif
F;1053-1055/Region: cell attachment (R-G-D) motif
F;1396-1398/Region: cell attachment (R-G-D) motif
F;1516-1744/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
F;1516-1627, 1628-1744/Region: duplication
F;1580-1586, 1691-1697/Disulfide bonds: #status predicted

Query Match 4.5%; Score 11; DB 2; Length 1744;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCLEEFRA 184
|||
Db 1675 SPGSCLEEFRA 1695

RESULT 18
collagen type IV alpha 2 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T13990
R;Yasochornsrikul, S.; Davis, W.J.; Cramer, G.; Kimbrell, D.A.; Dearolf, C.R.
submitted to the EMBL Data Library, July 1996
A;Description: Vikiing: identification and characterization of a novel type IV collagen
A;Reference number: Z17845
A;Accession: T13990
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1761 <YAS>
A;Cross-references: EMBL:U65431; NID:G2281290; PID:G2281291; PIDN:AAB64082.1
C;Genetics:
A;Gene: COLA2
A;Cross-references: FlyBase:FBgn0016075
C;Superfamily: collagen alpha 1(IV) chain

Query Match 4.5%; Score 11; DB 2; Length 1761;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCLEEFRA 184
|||
Db 1668 SPGSCLEEFRA 1678

RESULT 19
CGH02B
collagen alpha 2(IV) chain precursor - human
N;Alternate names: procollagen alpha 2(IV) chain
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
C;Accession: A32024; S00007; S02624; S00246; S17678; S16511; B32117; S16877; S00165; S39
R;Hostikka, S.L.; Tryggvason, K.
J. Biol. Chem. 263, 19488-19499, 1988
A;Title: The complete primary structure of the alpha2 chain of human type IV collagen an
A;Reference number: A32024; MUID:89066769; PMID:3198637
A;Accession: A32024
A;Molecule type: mRNA
A;Residues: 1-1712 <HOS1>
A;Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:G29550; PIDN:CAA29098.1; PI
R;Hostikka, S.L.; Kurkinen, M.; Tryggvason, K.
FEBS Lett. 216, 281-286, 1987

A;Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA re
ated region.
A;Reference number: S00007; MUID:87219158; PMID:3582677
A;Accession: S00007
A;Molecule type: mRNA
A;Residues: 1254-1398, 'V', 1400-1712 <HOS2>
A;Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:G29550; PIDN:CAA29098.1; PI
A;Note: 1399-1le was also found
R;Hostikka, S.L.; Tryggvason, K.
FEBS Lett. 224, 297-305, 1987
A;Title: Extensive structural differences between genes for the alpha(1) and alpha(2) ch
A;Reference number: S02624; MUID:88083553; PMID:2826228
A;Accession: S02624
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1347-1350; 1377-1383; 1426-1432; 1465-1471; 1529-1535; 1625-1630 <HOS3>
A;Note: complete nucleotide sequence not shown
R;Brazel, D.; Pollner, R.; Oberbaumer, I.; Kuehn, K.
Eur. J. Biochem. 172, 35-42, 1988
A;Title: Human basement membrane collagen (type IV): the amino acid sequence of the alph
A;Reference number: S00246; MUID:88151998; PMID:3345760
A;Accession: S00246
A;Molecule type: mRNA
A;Residues: 1-682, 'G', 684-1043 <BRA>
A;Cross-references: EMBL:X05562; NID:G30075; PIDN:CAA29076.1; PID:G30076
R;Oberbaumer, I.
submitted to the EMBL Data Library, June 1987
A;Reference number: S17678
A;Accession: S17678
A;Molecule type: mRNA
A;Residues: 1-470, 'P', 472-582, 'G', 684-1043 <OBE>
A;Cross-references: EMBL:X05562; NID:G30075; PIDN:CAA29076.1; PID:G30076
R;Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c
A;Reference number: S02738; MUID:89030632; PMID:2846280
A;Accession: S16911
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-33 <POE>
A;Cross-references: EMBL:J04217; EMBL:J05039; NID:G180759; PIDN:AAA53097.1; PID:G553233
R;Solinen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A;Reference number: A92850; MUID:89034231; PMID:3182844
A;Accession: B32117
A;Molecule type: DNA
A;Residues: 1-33 <SO11>
A;Cross-references: EMBL:J04217; EMBL:J05039; NID:G180759; PIDN:AAA53097.1; PID:G553233
R;Solinen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
A;Reference number: S16876; MUID:89340433; PMID:2701944
A;Accession: S16877
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-33 <SO12>
A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233; EMBL:J05039
A;Note: this sequence was submitted to the EMBL Data Library, October 1988
R;Siebold, B.; Qian, R.Q.; Glanville, R.W.; Hofmann, H.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 168, 569-575, 1987
A;Title: Construction of a model for the aggregation and cross-linking region (7S domain
is region.
A;Reference number: S00165; MUID:88029476; PMID:3117548
A;Accession: S00165
A;Molecule type: protein
A;Residues: 37-247 <S1E1>
A;Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-Gly
R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
A;Reference number: S39614; MUID:94038963; PMID:8223488
A;Accession: S39615

A:Molecule type: protein
A:Residues: 407-570 <EBL>
R:MacWhright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Pietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (B)
A:Reference number: S16910; MUID:84053346; PMID:6416291
A:Accession: S16912
A:Molecule type: protein
A:Residues: 490-492, 'X', 494-496; 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>
A:Experimental source: placenta
R:Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A:Title: Pepsin fragments of human placental basement-membrane collagens showing interru
A:Reference number: S16908; MUID:82005835; PMID:6792033
A:Accession: B58517
A:Molecule type: protein
A:Residues: 490-492, 'X', 494-501, 'P', 503-507; 952-957, 'X', 959-966, 'X', 968, 984-986, 'X', 988-
81-1185 <GLA>
R:Killen, P.D.; Francomano, C.A.; Yamada, Y.; Modi, W.S.; O'Brien, S.J.
Hum. Genet. 77, 318-324, 1987
A:Title: Partial structure of the human alpha-2(IV) collagen chain and chromosomal local
A:Reference number: S01450; MUID:88085168; PMID:3692475
A:Accession: S01450
A:Molecule type: mRNA
A:Residues: 1040, 'L', 1042-1398, 'V', 1400-1418, 'M', 1420-1635, 'V', 1637-1712 <KIL>
A:Cross-references: EMBL:M24766; NID:G537328; PIDN:AAA52043.1; PID:G537329
R:Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxy-term
A:Reference number: S02550; MUID:89005112; PMID:2844531
A:Accession: S02550
A:Molecule type: protein
A:Residues: 1480-1535; 1545-1614; 1617-1662, 'H', 1664-1700, 'G', 1705-1708; 1710-1712 <SIE2>
A:Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17
R:Myers, J.C.; Howard, P.S.; Jelen, A.M.; Dion, A.S.; Macarak, E.J.
J. Biol. Chem. 262, 9231-9238, 1987
A:Title: Duplication of type IV collagen COOH-terminal repeats and species-specific exp
A:Reference number: A27114; MUID:87250571; PMID:2439508
A:Accession: B27114
A:Molecule type: mRNA
A:Residues: 1486-1574, 'I', 1576-1712 <MYE>
A:Cross-references: EMBL:J02760; NID:G180425; PIDN:AAA58422.1; PID:G180426
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
A:Genetics:
A:Gene: GDB:COL4A2
A:Cross-references: GDB:119792; OMIM:120090
A:Map position: 13q34-13q34
A:Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1458/1; 1532/1; 1527/3 #status incomplete
A:Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with
C:Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:COH4B)
domains (with disulfide and desmosine cross-links), dimeric associations among trimer ca
rupted helical domain (with disulfide and desmosine cross-links).
C:Function:
A:Description: structural component of basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glycopro
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1712/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:29-57/Domain: amino-terminal nonhelical, NH1 <NH1>
F:58-1485/Region: interrupted helical
F:362-364/Region: cell attachment (R-G-D) motif
F:784-786/Region: cell attachment (R-G-D) motif
F:868-891/Region: cell attachment (R-G-D) motif
F:889-891/Region: cell attachment (R-G-D) motif
F:970-972/Region: cell attachment (R-G-D) motif
F:1069-1071/Region: cell attachment (R-G-D) motif
F:1228-1230/Region: cell attachment (R-G-D) motif
F:1452-1454/Region: cell attachment (R-G-D) motif
F:1486-1712/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1495-1593/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1603-1708/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:42, 47, 51, 53, 137, 483, 485/Disulfide bonds: interchain #status predicted

P:57,87,90,102,165,168,225,239,242/Binding site: carboxylate (Lys) (covalent) #status I
F:57/Modified site: 5-hydroxylysine (Lys) #status atypical
F:63,75,96,114,120,123,132,150,159,186,189,198,201,213,216,219,496,499,955,964,1103,1111
F:87,90,102,165,168,225,239,242/Modified site: 5-hydroxylysine (Lys) #status experimental
F:138/Binding site: carboxylate (Asn) (covalent) #status experimental
F:209/Modified site: 4-hydroxyproline (Pro) #status atypical
F:661-681/Disulfide bonds: #status predicted
F:1275/Binding site: carboxylate (Asn) (covalent) #status predicted
F:1590-1590,1537-1593/Disulfide bonds: (or 1504-1593, 1537-1590) #status experimental
F:1549-1555,1658-1665/Disulfide bonds: #status experimental
F:1612-1705,1646-1708/Disulfide bonds: (or 1612-1708, 1646-1705) #status experimental
Query Match 4.1%; Score 10; DB 1; Length 1712;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 128 PAIAIAVHSQ 137
Db 1596 PAIAIAVHSQ 1605
RESULT 20
A54121
collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)
N:Alternate names: collagen alpha 2(IV) chain homolog
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C:Accession: A54121; S44317
R:Exposito, J.F.; Suzuki, H.; Geourjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.
J. Biol. Chem. 269, 13167-13171, 1994
A:Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(IV)
A:Reference number: A54121; MUID:94230414; PMID:8175744
A:Accession: A54121
A:Molecule type: mRNA
A:Residues: 1-1747 <EXP>
A:Cross-references: EMBL:X76739; NID:G483606; PIDN:CAA54146.1; PID:G483607
C:Genetics:
A:Gene: COL4A1alpha
C:Superfamily: collagen alpha 1(IV) chain
Query Match 4.1%; Score 10; DB 2; Length 1747;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 117 PYISRCTVCE 126
Db 1622 PYISRCTVCE 1631
RESULT 21
I48303
collagen alpha 4(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 13-Aug-1999
C:Accession: I48303; S47279
R:Minier, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ
A:Reference number: A54979; MUID:95050957; PMID:7962065
A:Accession: I48303
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-312 <RES>
A:Cross-references: EMBL:Z35167; NID:G535199; PIDN:CAA84530.1; PID:G535200
C:Superfamily: collagen alpha 1(IV) chain
Query Match 3.7%; Score 9; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 233 ISRCQVCWK 241
Db 302 ISRCQVCWK 310

RESULT 22

A61228
collagen alpha 2(IV) chain precursor - rabbit (fragments)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
C:Accession: A61228
R;Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991
A:Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium
A:Reference number: A61228; MUID:92010685; PMID:1717398
A:Accession: A61228
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-775 <YAM>
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 3.7%; Score 9; DB 2; Length 775;
Best Local Similarity 100.0%; Pred No. 1;
Matches 9; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;

Qy 233 ISRCQVCMK 241

Db 765 ISRCQVCMK 773

RESULT 23

A33526
collagen alpha 2(IV) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: A33526; A24432; B25066; A24364; S19081; B29301; A24628; B32
E;Saus, J.; Quinones, S.; Mackrell, A.; Blumberg, B.; Muthukumaran, G.; Pihlajaniemi, T.
J. Biol. Chem. 264, 6318-6324, 1989
A:Title: The complete primary structure of mouse alpha-2(IV) collagen. Alignment with mouse alpha-1(IV) collagen.
A:Reference number: A33526; MUID:89197933; PMID:2703491
A:Accession: A33526
A:Molecule type: mRNA
A:Residues: 1-1707 <SAU>
A:Cross-references: EMBL:J04695; NID:9556298; PIDN:AAA50293.1; PID:9556299
R;Kurkinen, M.; Bernard, M.P.; Barlow, D.P.; Chow, L.T.
Nature 317, 177-179, 1985

A:Title: Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha-2(IV) chain
A:Reference number: A93367; MUID:85296379; PMID:3839908
A:Accession: A24432
A:Molecule type: mRNA

A:Residues: 967-1096 'G', 1098-1109 <KUL>
A:Cross-references: EMBL:X02896; NID:950263; PIDN:CAA26655.1; PID:9899326
A:Note: the authors translated the codon AAC for residue 964 as Lys
A:Accession: D24432
A:Molecule type: DNA

A:Residues: 964-1096 'G', 1098-1109 <KU2>
A:Cross-references: EMBL:X02899
R;Schwarz, U.; Schupp, D.; Oberbauer, I.; Glanville, R.W.; Deutzmann, R.; Timpl, R.; Eur. J. Biochem. 157, 49-56, 1986

A:Title: Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue
A:Reference number: A25066; MUID:86220192; PMID:3011432
A:Accession: A25066
A:Molecule type: mRNA

A:Residues: 970-1480 <SCI>
A:Cross-references: EMBL:X04647
A:Accession: B25066
A:Molecule type: protein

A:Residues: 979-1058;1065-1101;1105-1222;1226-1310;1326-1335;1351-1480 <SC2>
R;Vogeli, G.; Horn, E.; Carter, J.; Kaytes, P.S.
FEBS Lett. 206, 29-32, 1986

A:Title: Proposed alignment of helical interruptions in the two subunits of the basement membrane type IV collagen
A:Reference number: A24364; MUID:87005245; PMID:3758345
A:Accession: A24364
A:Molecule type: mRNA

A:Residues: 1041-1050, 'R', 1052-1170, 'S', 1172-1178, 'R', 1180-1240, 'E', 1242-1327, 'A', 1329-1

A:Cross-references: EMBL:X04410; NID:950240; PIDN:CAA27998.1; PID:9929678
R;Kaytes, P.S.; Theriault, N.Y.; Vogeli, G.
Gene 54, 141-146, 1987
A:Title: Homologies between the non-collagenous C-terminal (NC1) globular domains of the alpha-1(I) and alpha-2(I) collagen chains
A:Reference number: S19080; MUID:87277427; PMID:3609751
A:Accession: S19081
A:Molecule type: mRNA
A:Residues: 1466-1622 'H', 1624-1707 <KAI>
A:Cross-references: GB:X04410; NID:950240; PIDN:CAA27998.1; PID:9929678
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj J. Biol. Chem. 262, 8496-8499, 1987
A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV) and alpha-2(IV) collagen
A:Reference number: A94680; MUID:87250460; PMID:3597383
A:Accession: B29301
A:Molecule type: mRNA
A:Residues: 1481-1707 <KUR>
A:Cross-references: EMBL:M15833; NID:912284; PIDN:AAA37341.1; PID:9387116
R;Schwarz-Magdolen, U.; Oberbauer, I.; Kuehn, K.
FEBS Lett. 208, 203-207, 1986
A:Title: cDNA and protein sequence of the NC1 domain of the alpha-2-chain of collagen IV
A:Reference number: A24628; MUID:87054581; PMID:3780963
A:Accession: A24628
A:Molecule type: mRNA
A:Residues: 1480-1572, 'L', 1574-1622, 'H', 1624-1707 <SCH>
A:Cross-references: EMBL:X04647
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
J. Biol. Chem. 263, 19274-19277, 1988
A:Title: Head-to-head arrangement of murine type IV collagen genes.
A:Reference number: A92702; MUID:89066738; PMID:3198626
A:Accession: B32003
A:Molecule type: DNA
A:Residues: 1-33 <KA2>
A:Cross-references: EMBL:J04448; NID:912666; PIDN:AAA37438.1; PID:9126667
R;Burbello, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988

A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promoter
A:Reference number: A94220; MUID:89071759; PMID:3200851
A:Accession: B31786
A:Molecule type: DNA
A:Residues: 1-60 <BUR>
A:Cross-references: EMBL:M23333
R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A:Title: Subunit structure and assembly of the globular domain of basement-membrane collagen
A:Reference number: S17801; MUID:84132058; PMID:6698021
A:Accession: S19086
A:Molecule type: protein
A:Residues: 1475-1481, 'X', 1483-1487 <WEB>
C:Genetics:

A:Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1707/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F:29-171/Domain: 7S #status predicted <7SD>
F:58-1480/Domain: collagenous #status predicted <COL>
F:141-143/Region: cell attachment (R-G-D) motif
F:360-362/Region: cell attachment (R-G-D) motif
F:779-781/Region: cell attachment (R-G-D) motif
F:884-886/Region: cell attachment (R-G-D) motif
F:965-967/Region: cell attachment (R-G-D) motif
F:1223-1225/Region: cell attachment (R-G-D) motif
F:1447-1449/Region: cell attachment (R-G-D) motif
F:1481-1707/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1481-1589/Domain: repeat NC1 #status predicted <NC1>
F:1590-1707/Domain: repeat NC1 #status predicted <NC1>
F:42, 47, 51, 53, 481, 483/Disulfide bonds: interchain #status predicted
F:138, 1270/Binding site: carboxylate (Asn) (covalent) #status predicted
F:656-676, 1544-1550, 1653-1660/disulfide bonds: #status predicted
F:985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Modif
F:985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Modif
F:1009, 1012, 1018, 1021, 1024, 1037, 1040, 1043, 1046, 1052, 1058, 1070, 1098, 1110, 1128, 1140, 1149, 1
77, 1383, 1386, 1401, 1408, 1420, 1423, 1429, 1444, 1465, 1468, 1471, 1477/Modified site: hydroxypro

Query Match 3.3%; Score 8; DB 2; Length 281;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYISR 121
|||||
DB 91 ALEPYISR 98
CGHU6B

RESULT 26

N;collagen alpha 6(IV) chain precursor - human
N;alternate names: procollagen alpha 6(IV) chain
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 03-Oct-1995 #text_change 16-Jun-2000
C;Accession: A54122; A53404; B57079
R;Zhou, J.; Ding, M.; Zhao, Z.; Reeders, S.T.
J. Biol. Chem. 269, 13193-13199, 1994
A;Title: Complete primary structure of the sixth chain of human basement membrane collagen
A;Reference number: A54122; MUID:94230418; PMID:8175748
A;Accession: A54122
A;Molecule type: mRNA
A;Residues: 1-1691 <ZHO>
A;Cross-references: GB:U04845; NID:9496977; PIDN:AAA19569.1; PID:9496978
R;Ohashi, T.; Sugimoto, M.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 269, 7520-7526, 1994
A;Title: Identification of a new collagen IV chain, alpha6(IV), by cDNA isolation and as
A;Reference number: A53404; MUID:94171779; PMID:8125972
A;Accession: A53404
A;Molecule type: mRNA
A;Residues: 'MHPG', 6-169, 'M', 171-916, 'S', 918-1301, 1314-1355, 'A', 1357-1691 <OHO>
A;Cross-references: DDBJ:D21337; NID:9466537; PIDN:BAA04809.1; PID:9466538
R;Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggvason
Science 261, 1167-1169, 1993
A;Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm
A;Reference number: A57079; MUID:93361972; PMID:8556449
A;Accession: B57079
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546, 'G' <ZHZ>
A;Cross-references: GB:I22763
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A6
A;Cross-references: GDB:222775; OMIM:303631
A;Map position: Xq22-Xq22
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-li
C;Function:
A;Description: minor structural component of extracellular basement membrane
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1691/Product: collagen alpha 6(IV) chain #status predicted <MAT>
F;22-46/Domain: amino-terminal nonhelical, NC2 <NC2>
F;47-1463/Region: interrupted helical
F;1464-1691/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1473-1571/Domain: collagen IV carboxyl-terminal repeat <Crl>
F;1581-1687/Domain: collagen IV carboxyl-terminal repeat <Crl>
F;31,36,40,42,126,482,484,657/Disulfide bonds: interchain #status predicted
F;127/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1482-1568,1515-1571/Disulfide bonds: (or 1482-1571, 1515-1568) #status predicted
F;1527-1533,1636-1643/Disulfide bonds: #status predicted
F;1590-1684,1624-1687/Disulfide bonds: (or 1590-1687, 1624-1684) #status predicted

Query Match 3.3%; Score 8; DB 1; Length 1691;

Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
Db 1682 SRCQVCWK 1689

RESULT 27
A31893 collagen alpha 1(IV) chain precursor - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 21-Jul-2000
C:Accession: A31893; A26692; A19442; S00020
R:Blumberg, B.; Mackrell, A.J.; Fesseler, J.H.
J. Biol. Chem. 263, 18328-18337, 1988
A:Title: *Drosophila* basement membrane procollagen alpha-1(IV). II. Complete cDNA sequence
A:Reference number: A31893; MUID:89054012; PMID:3142875
A:Accession: A31893
A:Molecule type: mRNA
A:Residues: 1-1775 <BLU>
A:Cross-references: EMBL:W23704; NID:G157029; PIDN:AAA28404.1; PID:G157030
R:Blumberg, B.; Mackrell, A.J.; Olson, F.F.; Kurkinen, M.; Monson, J.M.; Natzle, J.E.; P.
J. Biol. Chem. 262, 5947-5950, 1987
A:Title: Basement membrane procollagen IV and its specialized carboxyl domain are conserved
A:Reference number: A26692; MUID:87194801; PMID:3106346
A:Accession: A26692
A:Molecule type: mRNA
A:Residues: 1065-1775 <BLU2>
A:Cross-references: EMBL:J02727
R:Monson, J.M.; Natzle, J.; Friedman, J.; McCarthy, B.J.
Proc. Natl. Acad. Sci. U.S.A. 79, 1761-1765, 1982
A:Title: Expression and novel structure of a collagen gene in *Drosophila*.
A:Reference number: A19442; MUID:82197577; PMID:6210912
A:Accession: A19442
A:Molecule type: DNA
A:Residues: 762-947, 'S', 949-996, 'T', 998-1230 <MON>
A:Cross-references: GB:J01074; EMBL:V00200; NID:G7736; PIDN:CAA23486.2; PID:G5777391
R:Cecchini, J.P.; Knibbe, B.; Mirre, C.; le Parco, Y.
Eur. J. Biochem. 165, 587-593, 1987
A:Title: Evidence for a type-IV-related collagen in *Drosophila melanogaster*. Evolutionary
A:Reference number: S00020; MUID:87246644; PMID:3109906
A:Accession: S00020
A:Molecule type: DNA
A:Residues: 1355-1356, 'K', 1358-1359, 'K', 1361-1372, 'I', 1374-1495, 'R', 1497-1506, 'RA', 1509,
A:Cross-references: EMBL:M28334
C:Genetics:
A:Gene: FlyBase:CG25C
A:Cross-references: FlyBase:FBgn0000299
A:Introns: 7/2; 23/3; 339/3; 505/2; 989/1; 1312/1; 1689/3
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-1775/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F:65-67/Region: cell attachment (R-G-D) motif
F:130-132/Region: cell attachment (R-G-D) motif
F:238-240/Region: cell attachment (R-G-D) motif
F:297-299/Region: cell attachment (R-G-D) motif
F:892-894/Region: cell attachment (R-G-D) motif
F:1075-1077/Region: cell attachment (R-G-D) motif
F:1173-1175/Region: cell attachment (R-G-D) motif
F:1225-1227/Region: cell attachment (R-G-D) motif
F:1545-1775/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>
F:1545-1655/Domain: repeat NCI #status predicted <NC11>
F:1656-1775/Domain: repeat NCI #status predicted <NC12>
F:72/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:570, 573/Disulfide bonds: interchain #status predicted
F:1611-1617, 1720-1727/Disulfide bonds: #status predicted

Query Match 3.3%; Score 8; DB 2; Length 1775;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
Db 1765 SRCQVCWK 1772

RESULT 28
S00999 superoxide dismutase (EC 1.15.1.1) (Fe) - *Azotobacter vinelandii* (fragment)
C:Species: *Azotobacter vinelandii*
C>Date: 15-Jul-1995 #sequence_revision 14-Nov-1997 #text_change 05-Mar-1999
C:Accession: S00999
R:Pagan, S.; Cohnaghi, R.; Palagi, A.; Negri, A.
FEBS Lett. 357, 79-82, 1995
A:Title: Purification and characterization of an iron superoxide dismutase from the nitr
A:Reference number: S00999; MUID:95094938; PMID:8001685
A:Accession: S00999
A:Molecule type: protein
A:Residues: 1-49 <PAG>
A:Experimental source: strain UW136
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Mn)
C:Keywords: homodimer; metalloprotein; oxidoreductase

Query Match 2.9%; Score 7; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 13 ALEPYIS 19

RESULT 29
C72639 hypothetical protein APE0547 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: C72639
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyr*
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: C72639
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <KAW>
A:Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79515.1; PID:d1043301; PID:G510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0547
C:Superfamily: *Aeropyrum pernix* hypothetical protein APE0547

Query Match 2.9%; Score 7; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 DLGTILGS 66
Db 52 DLGTILGS 58

RESULT 30
B42526 B3R protein - vaccinia virus (strain Copenhagen)
C:Species: vaccinia virus
A:Note: host *Homo sapiens* (man)
C>Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 08-Apr-1994
C:Accession: B42526
R:Johnson, G.P.
submitted to GenBank, June 1990
A:Reference number: A33172

Query Match 2.9%; Score 7; DB 2; Length 187;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIS 234
 |||||
 DB 61 ELEKIIS 67

RESULT 36
 A87671
 cytochrome c oxidase assembly protein, probable [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: A87671
 R:Kierman, W.C.; Fejdyblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: A87671
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-200 <STO>
 A:Cross-references: GB:AE005673; NID:g13425113; PIDN:AAK25365.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC3403
 C:Superfamily: cytochrome-c oxidase assembly protein COX11

Query Match 2.9%; Score 7; DB 2; Length 200;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115
 |||||
 DB 101 PITGRAL 107

RESULT 37
 S51097
 superoxide dismutase (EC 1.15.1.1) (Fe/Mn) - Methanobacterium thermoautotrophicum (strai
 C:Species: Methanobacterium thermoautotrophicum
 C>Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 20-Apr-2000
 C:Accession: S51097
 R:Maile, L.; Fischer, K.; Jenal, U.; Leisinger, T.
 submitted to the EMBL Data Library, July 1993
 A:Description: Molecular characterization of a superoxide dismutase gene from Methanoba
 A:Reference number: S51097
 A:Accession: S51097
 A:Molecule type: DNA
 A:Residues: 1-202 <MEI>
 A:Cross-references: EMBL:X74264; NID:g620124; PIDN:CAA52323.1; PID:g620125
 A:Experimental source: strain Marburg, DSM 2133
 C:Genetics:
 A:Gene: sod
 C:Function:

A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
 C:Superfamily: superoxide dismutase (Mn)
 C:Keywords: iron; manganese; metalloprotein; oxidoreductase
 F:30,78,164,168/Binding site: iron/manganese (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 202;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
 |||||
 DB 17 ALEPYIS 23

RESULT 38

JC4396
 superoxide dismutase (EC 1.15.1.1) (Fe/Mn) [validated] - Propionibacterium freudenreich
 C:Species: Propionibacterium freudenreichii subsp. shermanii
 C>Date: 20-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 20-Apr-2000
 C:Accession: JC4396; S41106
 R:Gabbianelli, R.; Battistoni, A.; Polizio, F.; Carri, M.T.; De Martino, A.; Meier, B.;
 Biochem. Biophys. Res. Commun. 216, 841-847, 1995
 A:Title: Metal uptake of recombinant cambialistic superoxide dismutase from Propionibac
 A:Reference number: JC4396; MUID:96074560; PMID:7488202
 A:Accession: JC4396
 A:Molecule type: DNA
 A:Residues: 1-202 <GAB>
 A:Cross-references: EMBL:X91650
 A:Experimental source: PZ3
 R:Meier, B.; Senn, A.P.; Schinina, M.E.; Barra, D.
 Eur. J. Biochem. 219, 463-468, 1994
 A:Title: In vivo incorporation of copper into the iron-exchangeable and manganese-exch
 A:Reference number: S41106; MUID:94139724; PMID:8307013
 A:Accession: S41106
 A:Molecule type: protein
 A:Residues: 2-202 <MEI>
 A:Experimental source: strain PZ3
 C:Genetics:
 A:Gene: sod
 C:Complex: homotetramer
 C:Function:

A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyge
 A:Note: can use iron or manganese as cofactor
 C:Superfamily: superoxide dismutase (Mn)
 C:Keywords: homotetramer; iron; manganese; metalloprotein; oxidoreductase
 F:2-202/Product: superoxide dismutase #status experimental <MAT>
 F:28,76,162,166/Binding site: iron/manganese (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 202;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
 |||||
 DB 15 ALEPYIS 21

RESULT 39
 E90174
 superoxide dismutase [Fe] (sod) [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
 C:Accession: E90174
 R:Jong, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Cha
 arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: E90174
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-211 <KUR>
 A:Cross-references: GB:AE006641; NID:g13813460; PIDN:AAK40652.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: sod
 C:Superfamily: superoxide dismutase (Mn)

Query Match 2.9%; Score 7; DB 2; Length 211;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
 |||||
 DB 21 ALEPYIS 27

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <NET>
A:Cross-references: GB:AE002359; GB:AE002098; NID:g7225525; PIDN:AAF40486.1; PID:g7225525
A:Experimental source: serogroup B, strain MCS8
C:Genetics:
A:Gene: NMB0007
C:Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 2.9%; Score 7; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GQDLGTL 64
|||
DB 63 GQDLGTL 69
|||

RESULT 43
DSBYN
superoxide dismutase (EC 1.15.1.1) (Mn) precursor [validated] - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YHR008c
C:Species: Saccharomyces cerevisiae
C:Date: 19-Feb-1994 #sequence_revision 17-Mar-1987 #text_change 23-Mar-2001
C:Accession: A00521; S46785; A90766
R:Marres, C.A.M.; Van Loon, A.P.G.M.; Oudshoorn, P.; Van Steeg, H.; Grivell, L.A.; Slater, E.; Blochem, 147, 153-161, 1985
A:Title: Nucleotide sequence analysis of the nuclear gene coding for manganese superoxide dismutase from *Saccharomyces cerevisiae*
A:Reference number: A91141; MUID:85127011; PMID:3882422
A:Accession: A00521
A:Molecule type: DNA
A:Residues: 1-233 <MAR>
A:Cross-references: EMBL:X02156; NID:g4513; PIDN:CAA26092.1; PID:g4514
R:Du, Z.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of *S. cerevisiae* cosmid L2825.
A:Reference number: S46774
A:Accession: S46785
A:Molecule type: DNA
A:Residues: 1-233 <DUZ>
A:Cross-references: EMBL:U10400; NID:g500701; PIDN:AAB68939.1; PID:g500704; GSPDB:GN0000C
R:Litlow, C.; Johansen, J.T.; Martin, B.M.; Svendsen, I.
Carlsberg Res. Commun. 47, 81-91, 1982
A:Title: The complete amino acid sequence of manganese-superoxide dismutase from *Saccharomyces cerevisiae*
A:Reference number: A90766
A:Accession: A90766
A:Molecule type: protein
A:Residues: 27-229 <DIT>
C:Genetics:
A:Gene: SGD:SOD2; MIPS:YHR008C
A:Cross-references: SGD:S0001050; MIPS:YHR008C
A:Map position: BR
A:Genome: nuclear
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Mn)
C:Keywords: homotetramer; manganese; metalloprotein; mitochondrial matrix; mitochondrion
F:1-26/Domain: transit peptide (mitochondrion) #status predicted <RNP>
F:27-229/Product: superoxide dismutase (Mn) #status experimental <MAT>
F:52,107,194,198/Binding site: manganese (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
|||
DB 39 ALEPYIS 45
|||

RESULT 44
H94035

hypothetical protein BH3088 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: H84035
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujita, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512592; PMID:11059132
 A:Accession: H84035
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-237 <STO>
 A:Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA806807.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3088
 C:Superfamily: Bacillus subtilis hypothetical protein yoaT

Query Match 2.9%; Score 7; DB 2; Length 237;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 VPLYSGF 45
 |||||
 Db 100 VPLYSGF 106

RESULT 45
 AB0520
 conserved hypothetical protein STY0161 [imported] - Salmonella enterica subsp. enterica
 C:Species: Salmonella enterica subsp. enterica serovar typhi
 A:Note: this species has also been called Salmonella typhi
 C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AB0520
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Fairy, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AB0520
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-247 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD01298.1; PID:G16501426; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY0161
 C:Superfamily: Escherichia coli hypothetical protein b0102

Query Match 2.9%; Score 7; DB 2; Length 247;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 WLASLNP 212
 |||||
 Db 157 WLASLNP 163

RESULT 46
 JG0179
 superoxide dismutase (EC 1.15.1.1) (Fe) - rice
 C:Species: Oryza sativa (rice)
 C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
 C:Accession: JG0179
 R:Kaminaka, H.; Morita, S.; Tokumoto, M.; Yokoyama, H.; Masumura, T.; Tanaka, K.
 Biosci. Biotechnol. Biochem. 63, 302-308, 1999
 A:Title: Molecular cloning and characterization of a cDNA for an iron-superoxide dismuta
 A:Reference number: JG0179; MUID:99208990; PMID:10192910
 A:Accession: JG0179
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-255 <KAM>

A:Cross-references: DDBJ:AB014056; NID:G4164148; PIDN:BAA37131.1; PID:G4164149
 C:Superfamily: superoxide dismutase (Mn)
 C:Keywords: iron; metalloprotein; oxidoreductase
 F:67,119,203,207/Binding site: Iron (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 255;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALPEYIS 120
 |||||
 Db 54 ALPEYIS 60

RESULT 47
 H82575
 3-deoxy-manno-octulosonate cytidylyltransferase XF2299 [imported] - Xylella fastidiosa (H82575)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: H82575
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: H82575
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-257 <SIM>
 A:Cross-references: GB:AE004041; GB:AE003849; NID:G9107453; PIDN:AAF85098.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
 as-Neto, E.; Docena, C.; El-Dostry, H.; Facincani, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000

A:Authors: Ferreira, J.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.C.R.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2299
 C:Superfamily: 3-deoxy-manno-octulosonate cytidylyltransferase

Query Match 2.9%; Score 7; DB 2; Length 257;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRFTH 74
 |||||
 Db 199 LQRFTH 205

RESULT 48
 WMS28
 complement factor D (EC 3.4.21.46) precursor - mouse
 N:Alternate names: adipocyte 28k proteinase; adipsin; C3 convertase activator; complemer
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 19-May-2000
 C:Accession: C25952; A00937; A26105
 R:Phillips, M.; Djian, P.; Green, H.
 J. Biol. Chem. 261, 10821-10827, 1986
 A:Title: The nucleotide sequence of three genes participating in the adipose differentia
 A:Reference number: A92553; MUID:86278164; PMID:3015943
 A:Accession: C25952
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-259 <PHI>
 A:Cross-references: GB:ML3386; NID:G192033; PIDN:AAA37262.1; PID:G387105

R;Cook, K.S.; Groves, D.L.; Min, H.Y.; Spiegelman, B.M.
Proc. Natl. Acad. Sci. U.S.A. 82, 6480-6484, 1985
A;Title: A developmentally regulated mRNA from 3T3 adipocytes encodes a novel serine pro
A;Reference number: A00937; MUID:86016726; PMID:3901003
A;Accession: A00937
A;Molecule type: mRNA
A;Residues: 1-259 <COO>
A;Cross-references: GB:M11768; NID:G202166; PIDN:AAA40486.1; PID:G202167
A;Experimental source: strain Swiss White
A;Note: Only one Ala is present in place of Ala-19 and Ala-20 in another equally abundan
R;Min, H.Y.; Spiegelman, B.M.
Nucleic Acids Res. 14, 8879-8892, 1986
A;Title: Adipin, the adipocyte serine protease: gene structure and control of expressio
A;Reference number: A26105; MUID:97066764; PMID:3024123
A;Accession: A26105
A;Molecule type: mRNA
A;Residues: 1-259 <MIN>
A;Cross-references: GB:X04673; NID:G493883; PIDN:CAA28378.1; PID:G581866
C;Comment: Human complement factor D is synthesized primarily in cells of the macrophage
plement activation. However, expression of the murine homolog is specific to adipose tis
C;Genetics:
A;Gene: 28K
A;Introns: 19/1; 71/2; 120/3; 206/3
C;Superfamily: trypsin; trypsin homology
C;Keywords: adipose tissue; alternative splicing; hydrolase; serine proteinase
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-25/Domain: propeptide #status predicted <PRO>
F;26-259/Product: adipin #status predicted <MAT>
F;26-249/Domain: trypsin homology <TRY>
F;51-67/149-215,180-196,205-230/Disulfide bonds: #status predicted
F;66,115,209/Active site: His, Asp, Ser #status predicted

Query Match 2.9%; Score 7; DB 1; Length 259;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSP 12
|||||
DB 206 RGDGSP 212

RESULT 49
A34476
collagen alpha 2(IV) chain - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C;Accession: A34476
R;Guo, X.; Kramer, J.M.
J. Biol. Chem. 264, 17574-17582, 1989
A;Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1
A;Reference number: A34476; MUID:90008929; PMID:2793871
A;Accession: A34476
A;Molecule type: DNA
A;Residues: 1-261 <GUO>
A;Cross-references: EMBL:J05066; NID:G156259; PIDN:AAA27989.1; PID:G156260
C;Genetics:
A;Gene: clb1
A;Map position: X
A;Introns: 77/1; 231/3
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F;1-30/Domain: collagenous (fragment) #status predicted <COL>
F;31-261/Domain: carboxyl-terminal nonhelical, NCL #status predicted <NCL1>
F;31-139/Domain: repeat NCL #status predicted <NCL1>
F;140-261/Domain: repeat NCL #status predicted <NCL2>
F;194-100,203-210/Disulfide bonds: #status predicted

Query Match 2.9%; Score 7; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLQRF 71
|||||

DB 80 GSCLQRF 86

RESULT 50
E69897
hypothetical protein yoar - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: E69897
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte:
C;Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch:
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle:
isch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue:
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Postecell:
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon:
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron:
Akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, J
A;Authors: Yoshikawa, H.P.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:96044033; PMID:9384377
A;Accession: E69897
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-264 <KUN>
A;Cross-references: GB:Z99114; GB:AL009126; NID:G2634230; PIDN:CAB13767.1; PID:G2634268
A;Experimental source: strain 168
C;Genetics:
A;Gene: yoar
C;Superfamily: Bacillus subtilis hypothetical protein yoar

Query Match 2.9%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
|||||
DB 103 VPLYSGF 109

RESULT 51
AD1561
B. subtilis yoar protein homolog lin1029 [imported] - Listeria innocua (strain Clip1126:
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AD1561
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker:
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; M:
ok, C.; Schlueter, T.; Smees, N.; Tierrez, A.; Vazquez-Boland, J.A.; Vose, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AD1561
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96260.1; PID:G16413488; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin1029
C;Superfamily: Bacillus subtilis hypothetical protein yoar

Query Match 2.9%; Score 7; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45

Db 103 VPLYSGF 109
|||||

RESULT 52

AE1204
B. subtilis YoaT protein homolog lmo1037 [imported] - Listeria monocytogenes (strain EGD-
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AE1204
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99115.1; PID:g16410439; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1037
C:Superfamily: Bacillus subtilis hypothetical protein yoaT

Query Match 2.9%; Score 7; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 VPLYSGF 45
|||||
Db 103 VPLYSGF 109

RESULT 53

AE6241
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
C:Accession: AE6241
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AE6141; MUID:21016719; PMID:11130712
A:Accession: AE6241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-343 <STO>
A:Cross-references: GB:AE005172; NID:g4874273; PIDN:AAD31338.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

C:Superfamily: Arabidopsis thaliana hypothetical protein F16G20.50
Query Match 2.9%; Score 7; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CPEGTVP 40
|||||
Db 55 CPEGTVP 61

RESULT 54

AE1204

C83577

hypothetical protein PA0549 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83577
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,
L.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C83577
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AE004491; GB:AE004091; NID:g9946412; PIDN:AAG03938.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0549

Query Match 2.9%; Score 7; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCLE 180
|||||
Db 115 SPGSCLE 121

RESULT 55

YS2699
hypothetical protein YDR191w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YD9346.03
C:Species: Saccharomyces cerevisiae
C:Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C:Accession: YS2699
R:Oliver, K.; Harris, D.
submitted to the EMBL Data Library, March 1995
A:Reference number: YS2699
A:Accession: YS2699
A:Molecule type: DNA
A:Residues: 1-370 <OLI>
A:Cross-references: EMBL:Z48784; NID:g755782; PIDN:CAA88705.1; PID:g755785; MIPS:YDR191w
A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:HST4
A:Cross-references: SGD:S0002599; MIPS:YDR191w
A:Map position: 4R

Query Match 2.9%; Score 7; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 KP1PSTV 224
|||||
Db 204 KP1PSTV 210

RESULT 56

YS7689
hypothetical protein YGR225w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G8541
C:Species: Saccharomyces cerevisiae
C:Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 19-Apr-2002
C:Accession: YS7689; S64549; S63905
R:van der Aart, Q.J.M.; Kleine, K.; Steensma, H.Y.
submitted to the EMBL Data Library, June 1995
A:Description: Sequence analysis of the 43 KB CRM1-YLM9-PET54-SMT1-PHO81-YHB4-PFK1 regic
A:Reference number: YS7689
A:Accession: YS7689
A:Molecule type: DNA
A:Residues: 1-409 <VAN>
A:Cross-references: EMBL:X87941; NID:G886908; PIDN:CAA61174.1; PID:G886918

A;Experimental source: strain S288C
R;van der Aart, Q.J.M.; Steensma, H.Y.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64541
A;Accession: S64549
A;Molecule type: DNA
A;Residues: 1-409 <VAV>
A;Cross-references: EMBL:Z73010; NID:gl323405; PIDN:CAA97253.1; PID:e243661; PID:gl32340
A;Experimental source: strain S288C
R;van der Aart, Q.J.M.; Kline, K.; Steensma, H.Y.
Yeast 12, 385-390, 1996
A;Title: Sequence analysis of the 43 kb CRM1-YLM3-PPT54-DIE2-SM11-PHO81-YHB4-PFK1 region
A;Reference number: S63896; MUID:96267763; PMID:9701610
A;Accession: S63905
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-409 <VAF>
A;Cross-references: EMBL:X87941; NID:G886908; PIDN:CAA61174.1; PID:G886918
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
C;Genetics:
A;Gene: SGD:AMA1
A;Cross-references: SGD:S0003457
A;Map position: 7R
A;Note: YGR225w

Query Match 2.9%; Score 7; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYS 94
|||||

DB 112 ASRNDYS 118

RESULT 57

S41607
atrolysin A (EC 3.4.24.1) - western diamondback rattlesnake (fragment)
N;Alternate names: hemorrhagic toxin a
C;Species: Crotalus atrox (western diamondback rattlesnake)
C;Date: 29-Sep-1994 #sequence_revision 13-Mar-1997 #text_change 09-Jun-2000
C;Accession: S41607
R;Hite, L.A.; Jia, L.G.; Bjarnason, J.B.; Fox, J.M.
Arch. Biochem. Biophys. 308, 182-191, 1994
A;Title: cDNA sequences for four snake venom metalloproteinases: structure, classification
A;Reference number: S41607; MUID:94145078; PMID:8311451
A;Accession: S41607

A;Status: preliminary; translation not shown

A;Molecule type: mRNA

A;Residues: 1-419 <HIT>

A;Cross-references: EMBL:U01234; NID:G402257; PID:G402258

C;Superfamily: mouse meltrin alpha; disintegrin homology

C;Keywords: hydrolase; metalloproteinase

F;209-291/Domain: disintegrin homology <DIS>

F;143/Active site: Glu #status predicted

Query Match 2.9%; Score 7; DB 2; Length 419;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 LFCNVND 83
|||||

DB 367 LFCNVND 373

RESULT 58

H89860
conserved hypothetical protein SA0804 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: H89860
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: H89860
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-438 <KUR>
A;Cross-references: GB:BA000018; PID:gl3700747; PIDN:BA842043.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0804
C;Superfamily: conserved integral membrane protein HP0758

Query Match 2.9%; Score 7; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSPA 13
|||||

DB 386 GDSGSPA 392

RESULT 59

T38239
hypothetical protein SPAC23C11.01 - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T38239

R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1995

A;Reference number: Z21781

A;Accession: T38239

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-441 <BRO>

A;Cross-references: EMBL:Z98559; PIDN:CA811154.1; GSPDB:GN00066; SPDB:SPAC23C11.01

A;Experimental source: strain 972h; cosmid c23C11

C;Genetics:

A;Gene: SPDB:SPAC23C11.01

A;Map position: 1

A;Introns: 88/3; 193/3; 255/3; 293/2

Query Match 2.9%; Score 7; DB 2; Length 441;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TGOALAS 174
|||||

DB 430 TGOALAS 436

RESULT 60

E70590
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) - Mycobacterium tuberculosis

N;Alternate names: aroA protein

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: E70590; A37807

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: E70590

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-450 <COL>

A;Cross-references: GB:Z95121; GB:AL123456; NID:G3261742; PIDN:CA808328.1; PID:G2072694

A;Experimental source: strain H37Rv

R;Garbe, T.; Jones, C.; Charles, I.; Dougan, G.; Young, D.

J. Bacteriol. 172, 6774-6782, 1990

A:Title: Cloning and characterization of the *aroA* gene from *Mycobacterium tuberculosis*.
A:Reference number: A37807; MUID:91072223; PMID:2123856
A:Accession: A37807
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-450 <GAR>
A:Cross-references: GB:M62708; MID:g149927; PIDN:AAA25356.1; PID:g149928
C:Genetics:
A:Gene: *aroA*
C:Superfamily: 3-phosphoshikimate 1-carboxyvinyltransferase; 3-phosphoshikimate 1-carboxyvinyltransferase
C:Keywords: transferase
F:15-417/Domain: 3-phosphoshikimate 1-carboxyvinyltransferase homology <PSK>

Query Match 2.9%; Score 7; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||||
DB 321 ALASPGS 327

RESULT 61
T30603
perlecan homolog 2L - Molluscum contagiosum virus 1
N:Alternate names: MC002L
C:Species: Molluscum contagiosum virus 1
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-May-2000
C:Accession: T30603
R:Senkevich, T.G.; Bugert, J.J.; Sieler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.
Science 273, 813-816, 1996
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re
A:Reference number: 220876; MUID:96325459; PMID:8670425
A:Accession: T30603
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-451 <SEN>
A:Cross-references: EMBL:U60315; PIDN:AAC55130.1
C:Genetics:
A:Note: MC002L

Query Match 2.9%; Score 7; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 TSAGSEG 167
|||||
DB 156 TSAGSEG 162

RESULT 62
S18804
collagen alpha 4(IV) chain - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 08-Nov-1996
C:Accession: S18804; S28834; E35167; S26673; D39419; A56630; S19076
R:Maruyama, M.; Kalluri, R.; Hudson, B.G.; Reeders, S.T.
J. Biol. Chem. 267, 1253-1258, 1992
A:Title: The alpha 4(IV) chain of basement membrane collagen. Isolation of cDNAs encoding
A:Reference number: S18804; MUID:92112769; PMID:1370461
A:Accession: S18804
A:Molecule type: mRNA
A:Residues: 1-453 <MAR>
A:Cross-references: GB:M77480
A:Accession: S20834
A:Molecule type: protein
A:Residues: 317, 'A', 319-322, 'S', 324-328 <MA2>
R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J. Biol. Chem. 265, 5466-5469, 1990
A:Title: Glomerular basement membranes. Identification of a fourth chain, alpha4, of type
A:Reference number: A35167; MUID:90202779; PMID:2318822
A:Accession: B35167
A:Molecule type: protein

A:Residues: 217-218, 'P', 220-242, 'X', 244-246 <GUN>
R:Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J. Biol. Chem. 262, 7874-7877, 1987
A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane c
A:Reference number: S18432; MUID:87222419; PMID:2438283
A:Accession: S20673
A:Molecule type: protein
A:Residues: 217-218, 'P', 220-233 <BUT>
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
J. Biol. Chem. 266, 15318-15324, 1991
A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll.
A:Reference number: A39419; MUID:91332055; PMID:1869555
A:Accession: D39419
A:Molecule type: protein
A:Residues: 217-237 <GU2>
R:Matsumura, H.; Michael, A.F.; Fish, A.J.; Butkowski, R.J.
Connect. Tissue Res. 28, 231-244, 1992
A:Title: Partial protein sequence of the globular domain of alpha 4(IV) collagen chain:
A:Reference number: A56630; MUID:93105615; PMID:1468209
A:Accession: A56630
A:Molecule type: protein
A:Residues: 217-218, 'P', 220-242, 'X', 244-256; 258-275, 'X', 277-278; 303-314; 391-396, 'X', 398-
A:Experimental source: kidney, basement membrane
A:Note: sequence modified after extraction from NCBI backbone
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; coiled coil; disulfide bond; duplication; extracellular m
F:1-222/Domain: collagenous (fragment) #status predicted <COL>
F:223-453/Domain: carboxyl-terminal nonhelical, NCI <NCI>
F:223-333; 334-453/Region: duplication
F:288-294; 397-404/Disulfide bonds: #status predicted

Query Match 2.9%; Score 7; DB 2; Length 453;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
|||||
DB 381 SPGSCLE 387

RESULT 63
D72344
DNA polymerase III, gamma and tau subunit - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Dec-2002
C:Accession: D72344
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72344
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-478 <ARN>
A:Cross-references: GB:AE001741; GB:AE000512; MID:g4981208; PIDN:RAD35768.1; PID:g498120
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0686
C:Superfamily: replication factor C large chain

Query Match 2.9%; Score 7; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 IISRCQV 238
|||||
DB 164 IISRCQV 170

RESULT 64

HMIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C:Species: influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C/Accession: A29971
R/Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A/Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A/Reference number: A94370; MUID:88101364; PMID:3336940
A/Accession: A29971
A/Molecule type: Genomic RNA
A/Cross-references: GB:M19056; NID:G324208
A/Note: the sequence in GenBank entry FLAHAPA, release 106, (PID:G324209) differs from B
C/Genetics:
A/Map position: segment 4
C/Superfamily: influenza virus hemagglutinin
C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
RESULT 65
HMIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C:Species: influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C/Accession: B29971
R/Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A/Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A/Reference number: A94370; MUID:88101364; PMID:3336940
A/Accession: B29971
A/Molecule type: Genomic RNA
A/Cross-references: GB:M19057; NID:G324210
A/Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:G324211) differs from B
C/Genetics:
A/Map position: segment 4
C/Superfamily: influenza virus hemagglutinin
C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
RESULT 66
HMIV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N/Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C:Species: influenza A virus
C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C/Accession: A27813
R/Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A/Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A/Reference number: A94363; MUID:87265458; PMID:2440178
A/Accession: A27813
A/Molecule type: Genomic RNA
A/Residues: 1-550 <KID>
A/Cross-references: GB:M16737; NID:G324081; PID:AAA43143.1; PID:G324082
C/Genetics:
A/Map position: segment 4
C/Superfamily: influenza virus hemagglutinin
C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
RESULT 67
HMIV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N/Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C/Accession: B27813
R/Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A/Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A/Reference number: A94363; MUID:87265458; PMID:2440178
A/Accession: B27813
A/Molecule type: Genomic RNA
A/Residues: 1-550 <KID>
A/Cross-references: GB:M16738; NID:G324083
A/Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C/Genetics:
A/Map position: segment 4
C/Superfamily: influenza virus hemagglutinin
C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
RESULT 68
HMIV33
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N/Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C:Accession: C27813
 R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
 Virolgy 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: C27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domain: transmembrane #status predicted <TM1>
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
 Db 358 SEGTGQA 364

RESULT 69
 HMI1989
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: D27813
 R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
 Virolgy 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: D27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domain: transmembrane #status predicted <TM1>
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
 Db 358 SEGTGQA 364

RESULT 70
 HMI21
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
 C:Accession: D27813
 R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
 Virolgy 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: E27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16741; NID:g324089
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domain: transmembrane #status predicted <TM1>
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
 Db 358 SEGTGQA 364

RESULT 71
 HMI198
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
 C:Accession: F27813
 R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
 Virolgy 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: F27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16742; NID:g324091
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domain: transmembrane #status predicted <TM1>
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
 Db 358 SEGTGQA 364

RESULT 72
 HMI15
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: G27813
 R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
 Virolgy 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: G27813
 A:Molecule type: genomic RNA

A;Residues: 1-550 <KID>
A;Cross-references: GB:M16743; NID:G324093; PIDN:AAA43149.1; PID:G324094
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;18,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
Db 358 SEGTGQA 364

RESULT 73
HmV86
hemagglutinin precursor - influenza A virus (strain A/Mem/6/66 [H3N2]) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jul-1999
C;Accession: A29245
R;Katz, J.M.; Webster, R.G.
Virology 165, 446-456, 1988
A;Title: Antigenic and structural characterization of multiple subpopulations of H3N2 in
A;Reference number: A29245; MUID:88306236; PMID:3407150
A;Accession: A29245
A;Molecule type: genomic RNA
A;Residues: 1-550 <KAT>
A;Cross-references: GB:M21649; NID:G324295; PIDN:AAA43275.1; PID:G324296
C;Genetics:
A;Gene: HA
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;18,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
Db 358 SEGTGQA 364

RESULT 74
JQ1153
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1153
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1153
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00929; NID:G221279; PIDN:BAA00769.1; PID:G221280

A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
Db 358 SEGTGQA 364

RESULT 75
JQ1154
hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1154
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1154
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00930; NID:G221273; PIDN:BAA00770.1; PID:G221274
A;Note: the authors translated the codon GGG for residue 218 as Glu
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
Db 358 SEGTGQA 364

RESULT 76
JQ1155
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1155
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1155
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00931; NID:G221277; PIDN:BAA00771.1; PID:G221278
A;Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 53
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;


```

Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
    |||||
Db 358 SEGTGQA 364

RESULT 77
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/231/77) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01156
R:Yasuda, J.; Shorridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491; PMID:1875195
A:Accession: J01156
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00932; NID:9221275; PIDN:BAA00772.1; PID:9221276
A>Note: the authors translated the codon TCG for residue 215 as Pro and GGA for residue
A:Note: residues 528-532 are not shown in this publication
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
    |||||
Db 358 SEGTGQA 364

RESULT 78
HMIVE1
hemagglutinin precursor - influenza A virus (strain A/equine/Uruguay/1/63 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: A34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: A34064
A:Molecule type: Genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24718; GB:J04336; NID:G324024; PIDN:AAA43114.1; PID:G324025
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
    |||||
Db 358 SEGTGQA 364

```

```

Db 373 SEGTGQA 379

RESULT 79
HMIVE3
hemagglutinin precursor - influenza A virus (strain A/equine/Tokyo/71 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: C34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: C34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24720; GB:J04336; NID:G324018; PIDN:AAA43111.1; PID:G324019
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
    |||||
Db 373 SEGTGQA 379

RESULT 80
HMIVE4
hemagglutinin precursor - influenza A virus (strain A/equine/Algiers/72 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: D34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: D34064
A:Molecule type: Genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24721; GB:J04336; NID:G323996; PIDN:AAA43100.1; PID:G323997
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
    |||||
Db 373 SEGTGQA 379

```

RESULT 81

HMIVE6
hemagglutinin precursor - influenza A virus (strain A/equine/New Market/76 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: E34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: E34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24722; GB:J04336; NID:G324010; PIDN:AAA43107.1; PID:G324011
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171

|||||

Db 373 SEGTGQA 379

RESULT 82

HMIVE6
hemagglutinin precursor - influenza A virus (strain A/equine/Fontainebleau/76 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: F34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: F34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24723; GB:J04336; NID:G323998; PIDN:AAA43101.1; PID:G323999
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171

|||||

Db 373 SEGTGQA 379

RESULT 83

HMIVE7
hemagglutinin precursor - influenza A virus (strain A/equine/Romania/80 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: G34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: G34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24724; GB:J04336; NID:G324014; PIDN:AAA43109.1; PID:G324015
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171

|||||

Db 373 SEGTGQA 379

RESULT 84

HMIVE8
hemagglutinin precursor - influenza A virus (strain A/equine/Santiago/1/85 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: H34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: H34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24725; GB:J04336; NID:G324016; PIDN:AAA43110.1; PID:G324017
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171

|||||

Db 373 SEGTGQA 379

RESULT 85

HMIV69
hemagglutinin precursor - influenza A virus (strain A/equine/Tennessee/5/85 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: J34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virolgy 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: J34064
A:Molecule type: Genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24726; GB:J04336; NID:g324020; PIDN:AAA43112.1; PID:g324021
C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TML>
F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
DB 373 SEGTGQA 379

RESULT 86

HMIVET
hemagglutinin precursor - influenza A virus (strain A/equine/Kentucky/2/86 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: A34065
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virolgy 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: A34065
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24727; GB:J04336; NID:g324000; PIDN:AAA43102.1; PID:g324001
C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TML>
F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
DB 373 SEGTGQA 379

RESULT 87

HMIVH

HMIVEE
hemagglutinin precursor - influenza A virus (strain A/equine/Kentucky/1/87 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: B34065
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virolgy 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: B34065
A:Molecule type: Genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24728; GB:J04336; NID:g324002; PIDN:AAA43103.1; PID:g324003
C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TML>
F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
DB 373 SEGTGQA 379

RESULT 88

S33703
hemagglutinin - influenza A virus H3N8
C:Species: influenza A virus H3N8, equine influenza virus
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
C:Accession: S33703
R:Binns, M.M.; Daly, J.M.; Chirnside, E.D.; Mumford, J.A.; Wood, J.M.; Richards, C.M.; E
Arch. Virol. 130, 33-43, 1993
A:Title: Genetic and antigenic analysis of an equine influenza H3 isolate from the 1989
A:Reference number: S33703; MUID:93277383; PMID:8503788
A:Accession: S33703
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-565 <BIN>
A:Cross-references: EMBL:X68437; NID:g312668; PIDN:CAA48482.1; PID:g312669
A:Note: the authors translated the codon ACC for residue 403 as Arg
C:Superfamily: influenza virus hemagglutinin

Query Match 2.9%; Score 7; DB 2; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
DB 373 SEGTGQA 379

RESULT 89

HMIVH
hemagglutinin precursor - influenza A virus
C:Species: influenza A virus
C>Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C:Accession: A93705; A93233; A04051; A93231; A94441
R:Both, G.W.; Sleight, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza
A:Reference number: A93705; MUID:81053696; PMID:6253883
A:Accession: A93705

A:Molecule type: genomic RNA
A:Residues: 1-566 <BOT>
A:Cross-references: GB:V01103
A:Experimental source: strain A/NT/60/68/29C
A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/68/29C
R:Dopheide, T.A.; Ward, C.W.
FEBS Lett. 110, 181-183, 1980
A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A:Reference number: A91276; MUID:8019105; PMID:6768586
A:Contents: annotation; disulfide bonds
R:Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from A/NT/60/68/29C
A:Reference number: A93233; MUID:81030852; PMID:7421990
A:Accession: A93233
A:Molecule type: genomic RNA
A:Residues: 1-24, 'S', '26', 'D', '28-159, 'G', '161-197, 'I', '199-241, 'L', '243-249 <GET>
A:Experimental source: strain X-31[H3]
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status experimental
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
DB 374 SEGTGQA 380

RESULT 90
HMI VHA
hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
A:Accession: A93231; A04051
R:Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers, W.
Nature 286, 771-776, 1980
A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/NT/60/68/29C and A/NT/68/29C
A:Reference number: A93231; MUID:80254693; PMID:7402351
A:Accession: A93231
A:Molecule type: genomic RNA
A:Residues: 1-566 <VAR>
A:Cross-references: GB:J02090; NID:G324131; PIDN:AAA43178.1; PID:G324132
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status predicted
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
DB 374 SEGTGQA 380

RESULT 91
HMI VHM
hemagglutinin precursor - influenza A virus (strain A/Wsm/102/72)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000

C:Accession: A94441; A04051
R:Sleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, 1977
A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of C/England/1968/33
A:Reference number: A94441
A:Accession: A94441
A:Molecule type: genomic RNA
A:Residues: 1-566 <SLB>
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status predicted
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
DB 374 SEGTGQA 380

RESULT 92
HMI V6
hemagglutinin precursor - influenza A virus (strain A/England/321/77)
C:Species: influenza A virus
C:Date: 05-Apr-1983 #sequence_revision 14-Nov-1983 #text_change 22-Oct-1999
A:Accession: B92790; A92979; A04052
R:Hauptmann, R.; Clarke, L.D.; Mountford, R.C.; Bachmayer, H.; Almond, J.W.
J. Gen. Virol. 64, 215-220, 1983
A:Title: Nucleotide sequence of the haemagglutinin gene of influenza virus A/England/321/77
A:Reference number: A92790; MUID:83110955; PMID:6822816
A:Accession: B92790
A:Molecule type: genomic RNA
A:Residues: 1-566 <HAU>
A:Cross-references: GB:X05907; NID:G60694; PIDN:CAA29337.1; PID:G60695
A:Note: the authors translated the codon GUU for residue 14 as Asn, GCC for residue 16 as Arg
R:Both, G.W.; Sleigh, M.J.
J. Virol. 39, 663-672, 1981
A:Title: Conservation and variation in the hemagglutinins of Hong Kong subtype influenza A virus
A:Reference number: A92979; MUID:82033259; PMID:6169840
A:Accession: A92979
A:Molecule type: genomic RNA
A:Residues: 17-31, 'G', '33-148, 'S', '150-158, 'S', '160-171, 'E', '173-175, 'K', '177-187, 'G', '189-201
A:Cross-references: GB:J02092; NID:G324139; PIDN:AAA43182.1; PID:G324140
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status predicted
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
DB 374 SEGTGQA 380

RESULT 93
HMI VDU
hemagglutinin precursor - influenza A virus (strain A/duck/Ukraine/63)
C:Species: influenza A virus
C:Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 28-May-1999
C:Accession: A04053

R;Fang, R.; Min Jou, W.; Huylebroeck, D.; Devos, R.; Fiers, W.
Cell 25, 315-323, 1981
A;Title: Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal virus
A;Reference number: A04053; MUID:82025542; PMID:6169439
A;Accession: A04053
A;Molecule type: genomic RNA
A;Residues: 1-566 <PAN>
A;Cross-references: GB:J02109; GB:J02108; NID:G60756; PIDN:CAA24271.1; PID:G60757
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;536-552/Domain: transmembrane #status predicted <TM1>
F;530-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status predicted
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
DB 374 SEGTGQA 380

RESULT 94
HMIVV
hemagglutinin precursor - influenza A virus (strain A/Victoria/3/75)
C;Species: influenza A virus
C;Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 28-May-1999
C;Accession: A07994; A04050; A92790
R;Min Jou, W.; Verhoeven, M.; Devos, R.; Saman, E.; Fang, R.; Huylebroeck, D.; Fiers, W.
Cell 19, 683-696, 1980
A;Title: Complete structure of the hemagglutinin gene from the human influenza A/Victoria
A;Reference number: A90794; MUID:80155186; PMID:6153390
A;Accession: A90794
A;Molecule type: genomic RNA
A;Residues: 1-567 <MIN>
A;Cross-references: GB:V01098; GB:J02172; GB:M55060; NID:G60784; PIDN:CAA24281.1; PID:G60784
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-345/Product: hemagglutinin HA1 #status predicted <HA1>
F;347-567/Product: hemagglutinin HA2 #status predicted <HA2>
F;537-553/Domain: transmembrane #status predicted <TM1>
F;31-483,69-294,81-93,156-490,298-322/Disulfide bonds: #status predicted
F;556,563,566/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 567;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
DB 375 SEGTGQA 381

RESULT 95
A45137
collagen alpha 4(IV) chain - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999
C;Accession: A45137
R;Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
A;Reference number: S28777; MUID:93054733; PMID:1429714
A;Accession: A45137
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA

A;Residues: 1-623 <KAM>
A;Experimental source: basement membrane
A;Note: sequence extracted from NCBI backbone (NCBIP:118549)
C;Superfamily: collagen alpha 1(IV) chain

Query Match 2.9%; Score 7; DB 2; Length 623;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPOSCLE 180
|||
DB 551 SPOSCLE 557

RESULT 96
A30347
exotoxin A precursor - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 24-Nov-1999
C;Accession: A30347
R;Gray, G.L.; Smith, D.H.; Baldridge, J.S.; Harkins, R.N.; Vasil, M.L.; Chen, E.Y.; Heyt
Proc. Natl. Acad. Sci. U.S.A. 81, 2645-2649, 1984
A;Title: Cloning, nucleotide sequence, and expression in Escherichia coli of the exotoxi
A;Reference number: A30347; MUID:84194063; PMID:6201861
A;Accession: A30347
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-638 <GRA>
A;Cross-references: GB:K01397; GB:M23348; NID:G151215; PIDN:AAB59097.1; PID:G151216
C;Superfamily: Pseudomonas aeruginosa exotoxin A
C;Keywords: exotoxin

Query Match 2.9%; Score 7; DB 2; Length 638;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||
DB 340 ALASPGS 346

RESULT 97
CB3503
exotoxin A precursor PA1148 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: CB3503
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: AB2950; MUID:20437337; PMID:10984043
A;Accession: CB3503
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-638 <STO>
A;Cross-references: GB:AE004544; GB:AE004091; NID:G9947060; PIDN:AA04537.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: toxA; PA1148
C;Superfamily: Pseudomonas aeruginosa exotoxin A

Query Match 2.9%; Score 7; DB 2; Length 638;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||
DB 340 ALASPGS 346

RESULT 98

I38755
transcription factor REST (version 2) - human (fragment)
N:Alternate names: neural-restrictive silencer factor; RE1-silencing transcription factor
C:Species: Homo sapiens (man)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999
C:Accession: I38755
R:Schonherr, C.J.; Anderson, D.J.
Science 267, 1360-1363, 1995
A:Title: The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple genes
A:Reference number: I38754; MUID:95176234; PMID:7871435
A:Accession: I38755
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-681 <RES>
A:Cross-references: EMBL:U13879; NID:G606947; PIDN:AAC50115.1; PID:G606948
C:Genetics:
A:Gene: GDB:REST; NRSF
A:Cross-references: GDB:702138
A:Map position: 4q12-4q12
C:Keywords: transcription regulation

Query Match 2.9%; Score 7; DB 2; Length 681;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSPA 13
|||||
Db 66 GDSGSPA 72

RESULT 99
S49228
sodium-dependent phosphate transporter - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 05-Nov-1999
C:Accession: S68972; S49228
R:Helms, C.; Murer, H.; McGivan, J.
Eur. J. Biochem. 228, 927-930, 1995
A:Title: Cloning, sequence analysis and expression of the cDNA encoding a sodium-dependent phosphate transporter
A:Reference number: S68972; MUID:95255303; PMID:7737195
A:Accession: S68972
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-693 <HE2>
A:Cross-references: EMBL:X81699; NID:G547483; PIDN:CAA57345.1; PID:G547484

Query Match 2.9%; Score 7; DB 2; Length 693;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||||
Db 472 ALASPGS 478

RESULT 100
AC0018
probable membrane protein YPO0142 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0018
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0018
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-715 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89005.1; PID:G15978247; GSPDB:GN00175
C:Genetics:

A:Gene: YPO0142
C:Superfamily: Escherichia coli hypothetical protein yrfF

Query Match 2.9%; Score 7; DB 2; Length 715;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TRGFVFT 23
|||||
Db 422 TRGFVFT 428

RESULT 101
T09395
envelope polyprotein - walleye dermal sarcoma virus
C:Species: walleye dermal sarcoma virus
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T09395
R:Petrooulos, C.J.
submitted to the EMBL Data Library, November 1997
A:Description: Appendix 2: Retroviral taxonomy, protein structure, sequences, and genetics
A:Reference number: Z16660
A:Accession: T09395
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: Genomic RNA
A:Residues: 1-1225 <PEI>
A:Cross-references: EMBL:AF033822; NID:G2801519; PID:G2801522
C:Genetics:
A:Gene: env
C:Superfamily: walleye dermal sarcoma virus envelope polyprotein

Query Match 2.9%; Score 7; DB 2; Length 1225;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 PMNWAPI 110
|||||
Db 967 PMNWAPI 973

RESULT 102
T13123
DNA replication primase protein - phase N15
N:Alternate names: protein gp37; replication protein
C:Species: phase N15
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Aug-1999
C:Accession: T13123; T13285
R:Hendrix, R.W.; Ravin, V.K.; Casjens, S.R.; Ford, M.E.; Ravin, N.V.; Smirnov, I.K.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z17603
A:Accession: T13123
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1324 <HEN>
A:Cross-references: EMBL:AF064539; NID:G3192683; PID:G3192713; PIDN:AAC19066.1
R:Vostrikhina, O.A.; Vostrov, A.A.; Rypchin, V.N.; Svarchevsky, A.N.
submitted to the EMBL Data Library, July 1996
A:Description: Characterization of the repA region of the Escherichia coli linear plasmid
A:Reference number: Z17650
A:Accession: T13285
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1207, 'VYQFTHVRWK', 1218, 'P', 1264, 'SK', 1267, 'TAIW', <VOS>
A:Cross-references: EMBL:U63085; NID:G2529379; PID:G2529380; PIDN:AAC48876.1
A:Experimental source: specific_host Escherichia coli
C:Genetics:
A:Gene: repA
A>Note: Gene 37

Query Match 2.9%; Score 7; DB 2; Length 1324;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SGFSFLF 49
 Db 105 SGFSFLF 111

RESULT 103
 T00333
 hypothetical protein KIAA0560 - human
 C/Species: Homo sapiens (man)
 C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 11-Jan-2002
 C/Accession: T00333
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
 DNA Res. 5, 31-39, 1998
 A/Title: Prediction of the coding sequences of unidentified human genes. IX. The completed
 A/Reference number: Z14086; MUID:98290545; PMID:9628851
 A/Accession: T00333
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-1421 <NAG>
 A/Cross-references: EMBL:AB011132; NID:dl185402; PIDN:BAA25486.1
 A/Experimental source: brain; clone HH1648
 C/Genetics:
 A/Note: KIAA0560

Query Match 2.9%; Score 7; DB 2; Length 1421;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 LEEFRAS 185
 Db 944 LEEFRAS 950

RESULT 104
 F86366
 protein F26F24.8 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: F86366
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: F86366
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1583 <STO>
 A/Cross-references: GB:AB005172; NID:9295691; PIDN:AAF86997.1; GSPDB:GN00141
 C/Genetics:
 A/Map position: 1

Query Match 2.9%; Score 7; DB 2; Length 1583;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKGRGD 8
 Db 773 LKGRGD 779

RESULT 105
 CGHUB
 collagen alpha 4(IV) chain precursor - human
 N/Alternate names: procollagen alpha 4(IV) chain
 C/Species: Homo sapiens (man)

C/Date: 06-Feb-1995 #sequence_revision 03-Oct-1995 #text_change 16-Jun-2000
 C/Accession: A55360; S36854; S28777
 R:Leinonen, A.; Mariyama, M.; Mochizuki, T.; Tryggvason, K.; Reenders, S.T.
 J. Biol. Chem. 269, 26172-26177, 1994
 A/Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. Comp
 A/Reference number: A55360; MUID:95014445; PMID:7523402
 A/Accession: A55360
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-1690 <LEI>
 A/Cross-references: GB:X81053; NID:9574805; PIDN:CAA56943.1; PID:9574806
 R:Sugimoto, M.; Ohashi, T.; Yoshiohka, H.; Matsuo, N.; Ninomiya, Y.
 FEBS Lett. 330, 122-128, 1993
 A/Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen cha
 A/Reference number: S36854; MUID:93374047; PMID:8365481
 A/Accession: S36854
 A/Molecule type: DNA; mRNA
 A/Residues: 1219-1658, 'FE', 1661-1690 <SUG>
 A/Cross-references: DDBJ:DL7391; NID:9440365; PIDN:BAA04214.1; PID:9457161
 A/Experimental source: whole eye
 R:Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
 J. Biol. Chem. 267, 23753-23759, 1992
 A/Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
 A/Reference number: S28777; MUID:93054733; PMID:1429714
 A/Accession: S28777
 A/Molecule type: DNA
 A/Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>
 A/Cross-references: GB:L01475; GB:L01476
 A/Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with the
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e
 ed and subsequently O-glycosylated.
 C/Genetics:
 A/Gene: GDB:COL4A4
 A/Cross-references: GDB:L32673; OMIM:120131
 A/Map position: 2q35-2q37
 A/Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
 A/Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C/Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3(
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
 er associations in the interrupted helical domain (with disulfide and desmosine cross-li
 C/Function:
 A/Description: minor structural component of extracellular basement membrane in kidney s
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxyl
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:59-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
 F:59-61/Domain: amino-terminal nonhelical, NH1 <NH1>
 F:62-1466/Region: interrupted helical
 F:94-96/Region: cell attachment (R-G-D) motif
 F:145-147/Region: cell attachment (R-G-D) motif
 F:189-191/Region: cell attachment (R-G-D) motif
 F:310-312/Region: cell attachment (R-G-D) motif
 F:724-726/Region: cell attachment (R-G-D) motif
 F:785-787/Region: cell attachment (R-G-D) motif
 F:989-991/Region: cell attachment (R-G-D) motif
 F:1212-1214/Region: cell attachment (R-G-D) motif
 F:1467-1690/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1579-1686/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:47,82,95,57,266,400,460,492,494,668,790,828,1095,1131,1294,1317,1375,1407/Disulfide bc
 F:142,669/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1480-1566/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
 F:1525-1531,1634-1641/Disulfide bonds: #status predicted
 F:1588-1683,1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 1690;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
 Db 1618 SPGSCLE 1624

RESULT 106

T29350
Hypothetical protein F01G12.5a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29350
R:Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F01G12.
A:Reference number: 220611
A:Accession: T29350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1758 <WUX>
A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a
A:Experimental source: strain Bristol N2; clone F01G12
C:Genetics:
A:Gene: CESP:F01G12.5a
A:Map position: X
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3
C:Superfamily: collagen alpha 1(IV) chain

Query Match 2.9%; Score 7; DB 2; Length 1758;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLORF 71
|||||
DB 1577 GSCLORF 1583

RESULT 107

T29351
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans
N:Alternate names: collagen alpha 2(IV) chain precursor clb-1
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29351
R:Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F01G12.
A:Reference number: 220611
A:Accession: T29351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1759 <WUX>
A:Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A:Experimental source: strain Bristol N2; clone F01G12
C:Genetics:
A:Gene: CESP:F01G12.5a
A:Map position: X
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3
C:Superfamily: collagen alpha 1(IV) chain

Query Match 2.9%; Score 7; DB 2; Length 1759;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLORF 71
|||||
DB 1578 GSCLORF 1584

RESULT 108

S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C:Species: Ascaris suum (pig roundworm)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: S16366
R:Petitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the part

A:Reference number: S16366; MUID:91340768; PMID:1714907
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <JBI>
A:Cross-references: GB:M67507; NID:G159648; PIDN:AAA18014.1; PID:G159649
C:Genetics:
A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfide
F1-26/Domain: signal sequence #status predicted <SIG>
F127-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F127-42/Domain: non-collagenous NH1 #status predicted <NH1>
F143-1529/Domain: collagenous #status predicted <COL>
F197-199/Region: cell attachment (R-G-D) motif
F1530-1763/Domain: carboxyl-terminal nonhelical, NCl #status predicted <NCL>
F1530-1638/Domain: repeat NCl #status predicted <NCL1>
F1633-1763/Domain: repeat NCl #status predicted <NCL2>
F131,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 2.9%; Score 7; DB 2; Length 1763;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMEPLFC 79
|||||
DB 1587 TMEPLFC 1593

RESULT 109

T08991
hypothetical protein F6G3.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 24-Nov-1999
C:Accession: T08991
R:Sevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.;
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216520
A:Accession: T08991
A:Molecule type: DNA
A:Residues: 1-1966 <BEV>
A:Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.180
A:Experimental source: cultivar Columbia; BAC clone F6G3
C:Genetics:
A:Gene: ATSP:F6G3.180
A:Map position: 4
A:Introns: 113/2; 652/3; 1112/3; 1220/2; 1720/1; 1869/2; 1934/2
C:Superfamily: Arabidopsis thaliana hypothetical protein F6G3.180

Query Match 2.9%; Score 7; DB 2; Length 1966;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTCQA 171
|||||
DB 1342 SEGTCQA 1348

RESULT 110

S02041
dystrophin, muscle - chicken
N:Alternate names: duchenne muscular dystrophy protein
C:Species: Gallus gallus (chicken)
C:Date: 07-Sep-1990 #sequence_revision 27-Jun-1994 #text_change 16-Jul-1999
C:Accession: S02041; S02013; S71487
R:Lemaire, C.; Heilig, R.; Mandel, J.L.
Nucleic Acids Res. 16, 11815-11816, 1988
A:Title: Nucleotide sequence of chicken dystrophin cDNA.
A:Reference number: S02041; MUID:89098331; PMID:3062582
A:Accession: S02041
A:Status: translation not shown
A:Molecule type: mRNA

A:Residues: 1-3660 <LEM>
A:Cross-references: EMBL:X13369; NID:963369; PIDN:CAA31746.1; PID:963370
A:Note: 1869-His, 1885-Arg, and sequences lacking 1171-Met were also found
R:Lemaire, C.; Heilig, R.; Mandel, J.L.
EMBO J. 7, 4157-4162, 1988
A:Title: The chicken dystrophin cDNA: striking conservation of the C-terminal coding and
A:Reference number: S02013; MUID:89210800; PMID:3072195
A:Accession: S02013
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3573; 'HA', 3576-3660 <LEM2>
R:Heilig, R.; Lemaire, C.; Mandel, J.L.
Nucleic Acids Res. 15, 9129-9142, 1987
A:Title: A 230kb cosmid walk in the Duchenne muscular dystrophy gene: detection of a cor
A:Reference number: S09071; MUID:88067745; PMID:2825128
A:Accession: S71487
A:Molecule type: DNA
A:Residues: 222-281 <HE1>
C:Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to the pl
C:Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dystro
C:Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystro
C:Keywords: actin binding; calmodulin binding; cytoskeleton; leucine zipper; membrane-as
F:18-233/Domain: alpha-actinin actin-binding domain homology <ACT>
F:253-327/Region: hinge
F:340-449/Domain: spectrin/dystrophin repeat homology <SP01>
F:450-558/Domain: spectrin/dystrophin repeat homology <SP02>
F:560-669/Domain: spectrin/dystrophin repeat homology <SP03>
F:670-719/Region: hinge
F:720-830/Domain: spectrin/dystrophin repeat homology <SP04>
F:838-936/Domain: spectrin/dystrophin repeat homology <SP05>
F:940-1047/Domain: spectrin/dystrophin repeat homology <SP06>
F:1049-1156/Domain: spectrin/dystrophin repeat homology <SP07>
F:1158-1265/Domain: spectrin/dystrophin repeat homology <SP08>
F:1267-1369/Domain: spectrin/dystrophin repeat homology <SP09>
F:1374-1479/Domain: spectrin/dystrophin repeat homology <SP10>
F:1480-1570/Domain: spectrin/dystrophin repeat homology <SP11>
F:1572-1678/Domain: spectrin/dystrophin repeat homology <SP12>
F:1680-1784/Domain: spectrin/dystrophin repeat homology <SP13>
F:1787-1877/Domain: spectrin/dystrophin repeat homology <SP14>
F:1878-1984/Domain: spectrin/dystrophin repeat homology <SP15>
F:1986-2103/Domain: spectrin/dystrophin repeat homology <SP16>
F:2105-2211/Domain: spectrin/dystrophin repeat homology <SP17>
F:2213-2319/Domain: spectrin/dystrophin repeat homology <SP18>
F:2323-2419/Domain: spectrin/dystrophin repeat homology <SP19>
F:2420-2467/Region: hinge
F:2468-2574/Domain: spectrin/dystrophin repeat homology <SP20>
F:2576-2683/Domain: spectrin/dystrophin repeat homology <SP21>
F:2685-2799/Domain: spectrin/dystrophin repeat homology <SP22>
F:2801-2928/Domain: spectrin/dystrophin repeat homology <SP23>
F:2930-3037/Domain: spectrin/dystrophin repeat homology <SP24>
F:3038-3075/Region: hinge
F:3082-3089/Domain: WW repeat homology <WW1>
F:3079-3357/Region: cysteine-rich
F:3481-3502/Region: leucine zipper motif
F:3547-3568/Region: leucine zipper motif

Query Match 2.9%; Score 7; DB 1; Length 3660;
Best Local Similarity 100.0%; Pred. No. 4.1e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 208 ASLNPER 214
Db 2284 ASLNPER 2290

RESULT 111
E39419
collagen alpha 5(IV) chain - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 19-Oct-1995
C:Accession: E39419
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
J. Biol. Chem. 266, 15318-15324, 1991

A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll
A:Reference number: A39419; MUID:91332055; PMID:1869555
A:Accession: E39419
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-20 <GUN>
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 2.5%; Score 6; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 53; Indels 0;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 23 TRHSQT 28
Db 15 TRHSQT 20

RESULT 112
T01664
envelope protein - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 17-Nov-2000
C:Accession: T01664
R:Scarselli, E.; Cerino, A.; Esposito, G.; Sillini, E.; Mondelli, M.U.; Traboni, C.
J. Virol. 69, 4407-4412, 1995
A:Title: Occurrence of antibodies reactive with more than one variant of the putative env
A:Reference number: Z14388; MUID:95287497; PMID:7539508
A:Accession: T01664
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-27 <SCA>
C:Cross-references: EMBL:X79669; NID:92276229; PIDN:CAA56117.1; PID:92276230
C:Genetics:
A:Gene: B2/NS1
C:Superfamily: hepatitis C virus genome polyprotein

Query Match 2.5%; Score 6; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 68; Indels 0;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 16 TTRGFV 21
Db 12 TTRGFV 17

RESULT 113
I39969
outer membrane protein A - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 19-Jul-1995 #sequence_revision 19-Jul-1996 #text_change 20-Aug-1999
C:Accession: I39969
R:Ikemura, H.; Takagi, H.; Inouye, M.
J. Biol. Chem. 262, 7859-7864, 1987
A:Title: Requirement of pro-sequence for the production of active subtilisin E in Escher
A:Reference number: I39969; MUID:87222417; PMID:3108260
A:Accession: I39969
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-35 <RES>
A:Cross-references: GB:M16639; NID:9143521; PIDN:AAA22743.1; PID:9143522
A:Experimental source: strain W168, substrain P379
A:Note: sequence was not translated in the genome sequence, reference number A69580
C:Genetics:
A:Gene: ompA
C:Superfamily: outer membrane protein A
C:Keywords: membrane protein

Query Match 2.5%; Score 6; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;

QY 129 AIAIV 134

R; Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

QY 230 EXIISR 235

Db 56 EXIISR 61

RESULT 119

hypothetical protein 65 - Cyanophora paradoxa cyanelle
N:Alternate names: protein ycf9
C:Species: Cyanelle Cyanophora paradoxa
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 21-Jan-2000
C/Accession: S14712; T06892; S10372
R:Evrad, J.L.; Weil, J.H.; Kuntz, M.
Plant Mol. Biol. 15, 779-781, 1990
A:Title: An ORF potentially encoding a 6.5 kDa hydrophobic protein in chloroplasts is a
A:Reference number: S14712; MUID:91346714; PMID:2129400
A:Accession: S14712
A:Molecule type: DNA
A:Residues: 1-65 <EVR>
A:Cross-references: EMBL:X51421; NID:g12550; PIDN:CAA35786.1; PID:g12551
A:Experimental source: strain LB555 UTEX
R:Stirwalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A:Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A:Reference number: Z15840
A:Accession: T06892
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-65 <STI>
A:Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81235.1; PID:g1016148
A:Experimental source: strain Pringsheim LB555
C:Genetics:
A:Gene: ycf9
A:Genome: cyanelle
C:Superfamily: conserved hypothetical protein ycf9
C:Keywords: cyanelle

Query Match 2.5%; Score 6; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 171 ALASPG 176
Db 26 ALASPG 31

RESULT 120

T06697
hypothetical protein T29H11.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C/Accession: T06697
R:Quetier, F.; Choisine, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artigou
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15793
A:Accession: T06697
A:Molecule type: DNA
A:Residues: 1-66 <QUE>
A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.30
A:Experimental source: cultivar Columbia, BAC Clone T29H11
C:Genetics:
A:Gene: ATSP:T29H11.30
A:Map position: 3
A:Introns: 52/1

Query Match 2.5%; Score 6; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 23 TRHSQT 28
Db 33 TRHSQT 38

RESULT 121

AB3399

hypothetical protein BME11176 [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C/Accession: AB3399
R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
M.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3399
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-72 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52357.1; PID:g17983153; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11176
A:Map position: 1

Query Match 2.5%; Score 6; DB 2; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 182 FRASPP 187
Db 25 FRASPP 30

RESULT 122

D82844
carbon storage regulator XP0125 [imported] - Xylella fastidiosa (strain 9a5C)
C:Species: Xylella fastidiosa
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C/Accession: D82844
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82844
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-76 <SIV>
A:Cross-references: GB:AE003866; GB:AE003849; NID:g9104906; PIDN:AAF82938.1; GSPDB:GN001
A:Experimental source: strain 9a5C
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, P.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir
as-Neto, E.; Docena, C.; El-Dorfi, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins F
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.F
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

RESULT 120

T06697
hypothetical protein T29H11.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C/Accession: T06697
R:Quetier, F.; Choisine, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artigou
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15793
A:Accession: T06697
A:Molecule type: DNA
A:Residues: 1-66 <QUE>
A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.30
A:Experimental source: cultivar Columbia, BAC Clone T29H11
C:Genetics:
A:Gene: ATSP:T29H11.30
A:Map position: 3
A:Introns: 52/1

Query Match 2.5%; Score 6; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 23 TRHSQT 28
Db 33 TRHSQT 38

RESULT 121

C:Superfamily: glycogen biosynthesis inhibitor

Query Match 2.5%; Score 6; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 6 RGDSGS 11
Db 69 RGDSGS 74

RESULT 123

T45103
H⁺-transporting two-sector ATPase (EC 3.6.3.14) chain K [imported] - Methanosarcina mazei
C:Species: Methanosarcina mazei
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Jun-2002
C:Accession: T45103
R:Ruppert, C.; Wimmers, S.; Muller, V.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z22913
A:Accession: T45103
A:Status: preliminary; translated from GB/EMBL/DDSB
A:Molecule type: DNA
A:Residues: 1-80 <RUP>
A:Cross-references: EMBL:U47274; PIDN:AAC06382.1
A:Experimental source: strain G01
C:Genetics:
A:Gene: ahA
C:Keywords: hydrolase

Query Match 2.5%; Score 6; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 129 AIAIAV 134
|||||
DB 22 AIAIAV 27

RESULT 124

AF2255
hypothetical protein asl13597 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2255
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-86 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB75296.1; PID:gl7132730; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asl13597

Query Match 2.5%; Score 6; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 227 GELEKI 232
|||||
DB 35 GELEKI 40

RESULT 125

AF1969
hypothetical protein asl1305 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF1969
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF1969
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-90 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAB73262.1; PID:gl7130652; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asl1305

Query Match 2.5%; Score 6; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 TTAIPS 33
|||||
DB 82 TTAIPS 87

RESULT 126

E83334
hypothetical protein PA2485 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83334
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim. ; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-92 <STO>
A:Cross-references: GB:AE004676; GB:AE004091; NID:g9948532; PIDN:AAG05873.1; GSPDB:GN00179
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2485

Query Match 2.5%; Score 6; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 FTSAGS 165
|||||
DB 28 FTSAGS 33

RESULT 127

JH0716
neuropeptide Y precursor - California sea hare
C:Species: Aplysia californica (California sea hare)
C>Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 21-Jul-2000
C:Accession: JH0716
R:Raybards, S.M.; Garcia, P.D.; Roberts, R.; Eliassen, J.C.; Owens, D.F.; Maltby, D.; Myle Neuron 9, 505-513, 1992
A:Title: Identification and molecular cloning of a neuropeptide Y homolog that produces a hyperlocomotor response in the California sea hare
A:Reference number: JH0716; MUID:92398969; PMID:1524828
A:Accession: JH0716
A:Molecule type: mRNA
A:Residues: 1-92 <RAJ>
A:Cross-references: GB:M98854; NID:gl55793; PIDN:AAA27772.1; PID:gl55794
A:Experimental source: abdominal ganglia
C:Function:
A:Description: neuropeptide inducing a number of behavioral effects including stimulatory effects on feeding, locomotion, and appetite; hormone; neuropeptide
C:Keywords: amidated carboxyl end; appetite; hormone; neuropeptide
F1-21/Domain: signal sequence #status predicted <SIG>
F12-61/Product: neuropeptide Y #status experimental <MAT>
F122-92/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F161/Modified site: amidated carboxyl end (Phe) (amide in mature form from following glycosylation)

Query Match 2.5%; Score 6; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GKRGDS 9

Db 62 GKEGDS 67
|||||

RESULT 128

T01876
hypothetical protein F8M12.13 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cross)
C/Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C/Accession: T01876
R/Madsen, C.; Graves, T.; Cotton, M.; Modde, T.
submitted to the EMBL Data Library, July 1998
A/Description: The sequence of A. thaliana F8M12.
A/Reference number: Z14450
A/Accession: T01876
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-93 <MAD>
A/Cross-references: EMBL:AF080118; NID:G3513725; PID:G3513729
A/Experimental source: cultivar Columbia
C/Genetics:
A/Map position: 4
A/Note: F8M12.13

Query Match 2.5%; Score 6; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKII 233
|||||
DB 20 ELEKII 25

RESULT 129

CCMP55
cytochrome c555 - Methylococcus capsulatus
C/Species: Methylococcus capsulatus
C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 03-Mar-2000
C/Accession: A23321; A05021
R/Amblar, R.P.; Dalton, H.; Meyer, T.E.; Bartsch, R.G.; Kamen, M.D.
Biochem. J. 233, 333-337, 1986
A/Title: The amino acid sequence of cytochrome c-555 from the methane-oxidizing bacterium
A/Reference number: A90328; MUID:86158741; PMID:3006666
A/Accession: A23321
A/Molecule type: protein
A/Residues: 1-96 <AMB>
A/Experimental source: strain Bath, NCIB 11132
C/Superfamily: cytochrome c6; cytochrome c6 homology
C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesis
F:8-79/Domain: cytochrome c6 homology <CYC>
F:19-22/Binding site: heme (Cys) (covalent) #status experimental
F:123,59/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 2.5%; Score 6; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 AGELEK 231
|||||
DB 89 AGELEK 94

RESULT 130

S30493
Sp1 protein - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C/Accession: S30493
R/Chestier, A.; Charnay, P.
DNA Seq. 2, 325-327, 1992
A/Title: Difference in the genomic organizations of the related transcription factors Sp1
A/Reference number: S30493; MUID:92338398; PMID:1633330
A/Accession: S30493

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-101 <CHE>
A/Cross-references: EMBL:X60136; NID:G54158; PIDN:CAA42721.1; PID:e38120; PID:gl334268

Query Match 2.5%; Score 6; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AGSEGT 168
|||||
DB 48 AGSEGT 53

RESULT 131

B90251
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C/Accession: B90251
R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001

A/Description: Sulfolobus solfataricus complete genome.

A/Reference number: A99139
A/Accession: B90251
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-103 <KUR>
A/Cross-references: GB:AE006641; NID:gl3814179; PIDN:AAK41265.1; GSPDB:GN00155
C/Genetics:
A/Gene: SSO0994

Query Match 2.5%; Score 6; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKII 233
|||||
DB 43 ELEKII 48

RESULT 132

T44890
hypothetical protein MLCB22.15C [imported] - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-May-2000
C/Accession: T44890
R/Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z22864
A/Accession: T44890

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 1-104 <PAR>
A/Cross-references: EMBL:Z98741; PIDN:CABL1380.1
A/Experimental source: cosmid B22
C/Genetics:
A/Note: MLCB22.15C
C/Superfamily: Mycobacterium leprae hypothetical protein MLCB22.15C

Query Match 2.5%; Score 6; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 ISRCTV 124
|||||
DB 79 ISRCTV 84

RESULT 133

E72599

probable formylmethanofuran dehydrogenase APE1261 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: E72599
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: E72599
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-105 <KAW>
 A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA080251.1; PID:G5104937
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1261
 C:Superfamily: Aeropyrum pernix probable formylmethanofuran dehydrogenase APE1261

Query Match 2.5%; Score 6; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPY 118
 |||||
 Db 12 RALEPY 17

RESULT 134
 H86901
 hypothetical protein ywJG [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C:Accession: H86901
 R:Boilotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis
 A:Reference number: A86625; MUID:21235186; PMID:111337471
 A:Accession: H86901
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-106 <STO>
 A:Cross-references: GB:AE005176; PID:G12725282; PIDN:AAK06314.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: ywJG

Query Match 2.5%; Score 6; DB 2; Length 106;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKII 233
 |||||
 Db 46 ELEKII 51

RESULT 135
 T42275
 hypothetical protein - phage SPPI
 C:Species: phage SPPI
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000
 C:Accession: T42275
 R:Alonso, J.C.; Luder, G.; Striege, A.C.; Chai, S.; Weise, F.; Trautner, T.A. Gene 204, 201-212, 1997
 A:Title: The complete nucleotide sequence and functional organization of Bacillus subtilis phage SPPI
 A:Reference number: 222137; MUID:98094274; PMID:9434185
 A:Accession: T42275
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-107 <ALO>
 A:Cross-references: EMBL:X97918; PIDN:CAA66585.1

Query Match 2.5%; Score 6; DB 2; Length 107;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 PLYSGF 45
 |||||
 Db 67 PLYSGF 72

RESULT 136
 B71524
 hypothetical protein CT357 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: B71524
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis serotype D
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: B71524
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-110 <ARN>
 A:Cross-references: GB:AE001309; GB:AE001273; NID:G3328777; PIDN:AA067953.1; PID:G332877
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: CT357

Query Match 2.5%; Score 6; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 LASPGS 177
 |||||
 Db 57 LASPGS 62

RESULT 137
 T46071
 hypothetical protein T18N14.130 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T46071
 R:Delseny, M.; Berger, C.; Cooke, R.; Greillett, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.; submitted to the Protein Sequence Database, December 1999
 A:Reference number: Z23013
 A:Accession: T46071
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-110
 A:Cross-references: EMBL:AL132968
 A:Experimental source: cultivar Columbia; BAC clone T18N14
 C:Genetics:
 A:Map position: 3
 A:Introns: 92/1
 A>Note: T18N14.130

Query Match 2.5%; Score 6; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIIS 234
 |||||
 Db 83 LEKIIIS 88

RESULT 138
 AF2540
 hypothetical protein all7609 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120b;
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AF2540
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi;

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AF2540
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-112 <KUR>
 A;Cross-references: GB:AP003602; PIDN:BAE77252.1; PID:g17134694; GSPDB:GN00181
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: all17609
 A;Genome: plasmid

Query Match 2.5%; Score 6; DB 2; Length 112;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 EEPFAS 185
 |||||
 DB 29 EEPFAS 34

RESULT 139
 C31769
 T-cell receptor delta-2 chain V region - human (fragment)
 C;Species: Homo sapiens (man)
 C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jan-2000
 C;Accession: C31769
 R;Loh, E.Y.; Cwirla, S.; Serafini, A.T.; Phillips, J.H.; Lanier, L.L.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9714-9718, 1988
 A;Title: Human T-cell-receptor delta chain: genomic organization, diversity, and expression
 A;Reference number: A94221; MUID:89071766; PMID:2974163
 A;Accession: C31769
 A;Molecule type: DNA
 A;Residues: 1-113 <LOH>
 A;Cross-references: GB:M23326; NID:G340877; PIDN:AAA61109.1; PID:G540457
 C;Genetics:
 A;Introns: 13/1
 C;Superfamily: immunoglobulin V region; immunoglobulin homology
 C;Keywords: T-cell receptor
 F;33-113/Domain: immunoglobulin homology <IMM>

Query Match 2.5%; Score 6; DB 2; Length 113;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GFSEFLF 49
 |||||
 DB 6 GFSEFLF 11

RESULT 140
 AH1784
 hypothetical protein lin2822 [imported] - Listeria innocua (strain Clip11262)
 C;Species: Listeria innocua
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C;Accession: AH1784
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maitournam, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, O.; C. Comparative Genomics of Listeria species.
 A;Title: Comparative Genomics of Listeria species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: AH1784
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-113 <GLA>
 A;Cross-references: GB:AL592022; PIDN:CAC98048.1; PID:g16415358; GSPDB:GN00178
 A;Experimental source: strain Clip11262
 C;Genetics:

A;Gene: lin2822

Query Match 2.5%; Score 6; DB 2; Length 113;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
 |||||
 DB 90 EGTVP 95

RESULT 141
 B72648
 hypothetical protein APE0616 - Aeropyrum pernix (strain K1)
 C;Species: Aeropyrum pernix
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C;Accession: B72648
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yanazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
 A;Reference number: A72450; MUID:99310339; PMID:10382966
 A;Accession: B72648
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-115 <KAW>
 A;Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79596.1; PID:d1043372; PID:G5104188
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE0616

Query Match 2.5%; Score 6; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPG 176
 |||||
 DB 85 ALASPG 90

RESULT 142
 T49363
 hypothetical protein B1D1.170 [imported] - Neurospora crassa
 C;Species: Neurospora crassa
 C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
 C;Accession: T49363
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
 A;Reference number: Z25022
 A;Accession: T49363
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-119 <SCH>
 A;Cross-references: EMBL:AL555927; GSPDB:GN00116; NCSP:B1D1.170
 A;Experimental source: BAC clone B1D1; strain OR74A
 C;Genetics:
 A;Gene: NCSP:B1D1.170
 A;Map position: 6
 C;Superfamily: Neurospora crassa hypothetical protein B1D1.170

Query Match 2.5%; Score 6; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SCPEGT 38
 |||||
 DB 11 SCPEGT 16

RESULT 143
 S10914
 probable phosphoesterase (EC 3.1.1.-) - Synecococcus sp. (PCC 6301)
 N;Alternate names: hypothetical protein 1 (16S rRNA 5' region)

C;Species: *Synechococcus* sp.
 C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 24-Sep-1999
 C;Accession: S10914
 R;Kumano, M.; Tomioka, N.; Shinozaki, K.; Sugliura, M.
 Mol. Gen. Genet. 202, 173-178, 1986
 A;Title: Analysis of the promoter region in the rna operon from a blue-green alga, *Anacystis*
 A;Reference number: S07311
 A;Accession: S10914
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-119 <KUM>
 A;Cross-references: EMBL:X03538; NID:G38918; PIDN:CAA7241.1; PID:G38921
 C;Comment: This sequence has motifs characteristic of a variety of phosphoesterases.
 C;Superfamily: unassigned probable phosphoesterases; phosphoesterase core homology
 C;Keywords: hydrolase
 F;4-82/Domain: phosphoesterase core homology <PEC>

Query Match 2.5%; Score 6; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIA 133
 |||||
 Db 73 PAIAIA 78

RESULT 144
 B95415
 C;Species: *Sinorhizobium meliloti*
 C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C;Accession: B95415
 R;Barnett, M.J.; Fisher, R.P.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
 proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A;Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
 A;Reference number: A95462; NUID:21396509; PMID:11481432
 A;Accession: B95415
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-126 <KUR>
 A;Cross-references: GB:AE006469; PIDN:AAK65884.1; PID:G14524393; GSPDB:GN00165
 A;Experimental source: Strain 1021, megaplasmid pSYMA
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Fuhrer, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A;Reference number: A96039; NUID:21368234; PMID:11474104
 A;Contents: annotation
 C;Genetics:
 A;Gene: SMA2277
 A;Genome: plasmid

Query Match 2.5%; Score 6; DB 2; Length 126;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SQTAT 31
 |||||
 Db 12 SQTAT 17

RESULT 145
 A72388
 C;Species: *Thermotoga maritima*
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C;Accession: A72388
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.
 Nature 399, 323-329, 1999
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se
 A;Reference number: A72200; NUID:99287316; PMID:10360571
 A;Accession: A72388
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-127 <ARN>
 A;Cross-references: GB:AE001715; GB:AE000512; NID:G4980839; PIDN:AAD35426.1; PID:G49808
 A;Experimental source: strain MSB8
 C;Genetics:
 A;Gene: TM0339
 C;Superfamily: *Thermotoga maritima* hypothetical protein TM0339

Query Match 2.5%; Score 6; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 FRASPF 187
 |||||
 Db 30 FRASPF 35

RESULT 146
 S49637
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 19-Apr-2002
 C;Accession: S49637
 R;Gentiles, S.; Bowman, S.
 submitted to the EMBL Data Library, November 1994
 A;Reference number: S49637
 A;Accession: S49637
 A;Molecule type: DNA
 A;Residues: 1-128 <GEN>
 A;Cross-references: EMBL:Z46660; NID:G575702; PID:G575713; GSPDB:GN00013; MIPS:YML090W
 C;Genetics:
 A;Gene: MIPS:YML090W
 A;Cross-references: SGD:S0004555
 A;Map position: 13L
 C;Superfamily: *Saccharomyces cerevisiae* probable membrane protein YML090W

Query Match 2.5%; Score 6; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 PSFLFV 50
 |||||
 Db 4 PSFLFV 9

RESULT 147
 A81062
 C;Species: *Neisseria meningitidis*
 C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C;Accession: A81062
 R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.;
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vi
 A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A;Reference number: A81000; NUID:20175755; PMID:10710307
 A;Accession: A81062
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-134 <RET>
 A;Cross-references: GB:AE000512; GB:AE002098; NID:G7226866; PIDN:AAF41971.1; PID:G72268
 A;Experimental source: serogroup B, strain MC58

C;Genetics:
A;Gene: NMB1619

Query Match 2.5%; Score 6; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DSGSPA 13
|||||
Db 31 DSGSPA 36

RESULT 148

G81807
hypothetical protein NMA1818 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: G81807
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: G81807
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-134 <PAR>
A;Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85043.1; PID:g738045
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA1818

Query Match 2.5%; Score 6; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DSGSPA 13
|||||
Db 31 DSGSPA 36

RESULT 149

C83908
cytochrome aa3 quinol oxidase subunit IV qoxD [imported] - Bacillus halodurans (strain C
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: C83908
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: C83908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-136 <STO>
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05786.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: qoxD

Query Match 2.5%; Score 6; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAY 134
|||||
Db 85 AIAIAY 90

RESULT 150

F70657
hypothetical protein Rv2530c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: F70657
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Comor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70657
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-139 <COL>
A;Cross-references: GB:Z83863; GB:AL123456; NID:g3261685; PIDN:CAB06178.1; PID:e290885;
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv2530c

Query Match 2.5%; Score 6; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIA 133
|||||
Db 62 PAIAIA 67

Search completed: April 5, 2004, 07:39:26
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:07:33 ; Search time 17 seconds
(without alignments)
747.360 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRGSGSPATWTRGF.....KAGELEKIISRCQVCMKKRH 244

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 244 | 100.0 | 1670 | 1 CA34 HUMAN | Q01955 homo sapien |
| 2 | 39 | 16.0 | 471 | 1 CA34 BOVIN | Q28084 bos taurus |
| 3 | 17 | 7.0 | 754 | 1 CA54 CANFA | Q28247 canis famil |
| 4 | 17 | 7.0 | 1669 | 1 CA14 HUMAN | P02462 homo sapien |
| 5 | 17 | 7.0 | 1669 | 1 CA14 MOUSE | P02463 mus musculus |
| 6 | 17 | 7.0 | 1685 | 1 CA34 HUMAN | P29400 homo sapien |
| 7 | 11 | 4.5 | 1758 | 1 CA14 CAEEL | P17139 caenorhabdi |
| 8 | 10 | 4.1 | 1712 | 1 CA24 HUMAN | P08572 homo sapien |
| 9 | 9 | 3.7 | 1707 | 1 CA24 MOUSE | P08122 mus musculus |
| 10 | 8 | 3.3 | 200 | 1 SODM AGABI | Q94856 agarius bi |
| 11 | 8 | 3.3 | 281 | 1 SODF EACSU | Q35023 bacillus su |
| 12 | 8 | 3.3 | 1691 | 1 CA54 HUMAN | Q14031 homo sapien |
| 13 | 8 | 3.3 | 1775 | 1 CA14 DROME | P08120 drosophila |
| 14 | 7 | 2.9 | 124 | 1 VB03 VACCC | P21000 vaccinia vi |
| 15 | 7 | 2.9 | 137 | 1 RUVK OCEIH | Q89PT1 oceanobacil |
| 16 | 7 | 2.9 | 151 | 1 RNB HSV2H | P89479 herpes simp |
| 17 | 7 | 2.9 | 183 | 1 NUPM NEUCR | P21976 neurospora |
| 18 | 7 | 2.9 | 201 | 1 SODM PROFR | P80293 propionibac |
| 19 | 7 | 2.9 | 202 | 1 SODF METTM | Q60036 methanobact |
| 20 | 7 | 2.9 | 210 | 1 SODF SULAC | Q08713 sulfobac |
| 21 | 7 | 2.9 | 210 | 1 SODF SULSO | P08057 sulfobac |
| 22 | 7 | 2.9 | 211 | 1 SODF ACTAM | Q9P913 acidianus a |
| 23 | 7 | 2.9 | 211 | 1 SODF PYRAE | Q93724 pyrobaculum |
| 24 | 7 | 2.9 | 233 | 1 SODM YEAST | P00447 saccharomyc |
| 25 | 7 | 2.9 | 245 | 1 SODM NEUCR | Q9Y783 neurospora |
| 26 | 7 | 2.9 | 247 | 1 YACF SALTU | Q8XEX8 salmonella |
| 27 | 7 | 2.9 | 259 | 1 CFAD MOUSE | P03953 mus musculus |
| 28 | 7 | 2.9 | 370 | 1 HST4 YEAST | P53688 saccharomyc |
| 29 | 7 | 2.9 | 409 | 1 YG4S YEAST | P50082 saccharomyc |
| 30 | 7 | 2.9 | 441 | 1 YDM1 SCHPO | O13909 schizosacch |
| 31 | 7 | 2.9 | 450 | 1 AROA MYCTU | P22487 mycobacteri |
| 32 | 7 | 2.9 | 453 | 1 CA44 BOVIN | Q29442 bos taurus |
| 33 | 7 | 2.9 | 488 | 1 GATE_RALSO | Q8Y3C6 ralstonia s |

| | | |
|-----|--------------|--------------------|
| 550 | HEMA_IABAN | P03441 influenza a |
| 551 | HEMA_IADH1 | P12582 influenza a |
| 552 | HEMA_IADH2 | P12583 influenza a |
| 553 | HEMA_IADH3 | P12584 influenza a |
| 554 | HEMA_IADH4 | P12585 influenza a |
| 555 | HEMA_IADH5 | P12586 influenza a |
| 556 | HEMA_IADH6 | P12587 influenza a |
| 557 | HEMA_IADH7 | P12588 influenza a |
| 558 | HEMA_IADH8 | P12589 influenza a |
| 559 | HEMA_IADH9 | P12590 influenza a |
| 560 | HEMA_IADH10 | P12591 influenza a |
| 561 | HEMA_IADH11 | P12592 influenza a |
| 562 | HEMA_IADH12 | P12593 influenza a |
| 563 | HEMA_IADH13 | P12594 influenza a |
| 564 | HEMA_IADH14 | P12595 influenza a |
| 565 | HEMA_IADH15 | P12596 influenza a |
| 566 | HEMA_IADH16 | P12597 influenza a |
| 567 | HEMA_IADH17 | P12598 influenza a |
| 568 | HEMA_IADH18 | P12599 influenza a |
| 569 | HEMA_IADH19 | P12600 influenza a |
| 570 | HEMA_IADH20 | P12601 influenza a |
| 571 | HEMA_IADH21 | P12602 influenza a |
| 572 | HEMA_IADH22 | P12603 influenza a |
| 573 | HEMA_IADH23 | P12604 influenza a |
| 574 | HEMA_IADH24 | P12605 influenza a |
| 575 | HEMA_IADH25 | P12606 influenza a |
| 576 | HEMA_IADH26 | P12607 influenza a |
| 577 | HEMA_IADH27 | P12608 influenza a |
| 578 | HEMA_IADH28 | P12609 influenza a |
| 579 | HEMA_IADH29 | P12610 influenza a |
| 580 | HEMA_IADH30 | P12611 influenza a |
| 581 | HEMA_IADH31 | P12612 influenza a |
| 582 | HEMA_IADH32 | P12613 influenza a |
| 583 | HEMA_IADH33 | P12614 influenza a |
| 584 | HEMA_IADH34 | P12615 influenza a |
| 585 | HEMA_IADH35 | P12616 influenza a |
| 586 | HEMA_IADH36 | P12617 influenza a |
| 587 | HEMA_IADH37 | P12618 influenza a |
| 588 | HEMA_IADH38 | P12619 influenza a |
| 589 | HEMA_IADH39 | P12620 influenza a |
| 590 | HEMA_IADH40 | P12621 influenza a |
| 591 | HEMA_IADH41 | P12622 influenza a |
| 592 | HEMA_IADH42 | P12623 influenza a |
| 593 | HEMA_IADH43 | P12624 influenza a |
| 594 | HEMA_IADH44 | P12625 influenza a |
| 595 | HEMA_IADH45 | P12626 influenza a |
| 596 | HEMA_IADH46 | P12627 influenza a |
| 597 | HEMA_IADH47 | P12628 influenza a |
| 598 | HEMA_IADH48 | P12629 influenza a |
| 599 | HEMA_IADH49 | P12630 influenza a |
| 600 | HEMA_IADH50 | P12631 influenza a |
| 601 | HEMA_IADH51 | P12632 influenza a |
| 602 | HEMA_IADH52 | P12633 influenza a |
| 603 | HEMA_IADH53 | P12634 influenza a |
| 604 | HEMA_IADH54 | P12635 influenza a |
| 605 | HEMA_IADH55 | P12636 influenza a |
| 606 | HEMA_IADH56 | P12637 influenza a |
| 607 | HEMA_IADH57 | P12638 influenza a |
| 608 | HEMA_IADH58 | P12639 influenza a |
| 609 | HEMA_IADH59 | P12640 influenza a |
| 610 | HEMA_IADH60 | P12641 influenza a |
| 611 | HEMA_IADH61 | P12642 influenza a |
| 612 | HEMA_IADH62 | P12643 influenza a |
| 613 | HEMA_IADH63 | P12644 influenza a |
| 614 | HEMA_IADH64 | P12645 influenza a |
| 615 | HEMA_IADH65 | P12646 influenza a |
| 616 | HEMA_IADH66 | P12647 influenza a |
| 617 | HEMA_IADH67 | P12648 influenza a |
| 618 | HEMA_IADH68 | P12649 influenza a |
| 619 | HEMA_IADH69 | P12650 influenza a |
| 620 | HEMA_IADH70 | P12651 influenza a |
| 621 | HEMA_IADH71 | P12652 influenza a |
| 622 | HEMA_IADH72 | P12653 influenza a |
| 623 | HEMA_IADH73 | P12654 influenza a |
| 624 | HEMA_IADH74 | P12655 influenza a |
| 625 | HEMA_IADH75 | P12656 influenza a |
| 626 | HEMA_IADH76 | P12657 influenza a |
| 627 | HEMA_IADH77 | P12658 influenza a |
| 628 | HEMA_IADH78 | P12659 influenza a |
| 629 | HEMA_IADH79 | P12660 influenza a |
| 630 | HEMA_IADH80 | P12661 influenza a |
| 631 | HEMA_IADH81 | P12662 influenza a |
| 632 | HEMA_IADH82 | P12663 influenza a |
| 633 | HEMA_IADH83 | P12664 influenza a |
| 634 | HEMA_IADH84 | P12665 influenza a |
| 635 | HEMA_IADH85 | P12666 influenza a |
| 636 | HEMA_IADH86 | P12667 influenza a |
| 637 | HEMA_IADH87 | P12668 influenza a |
| 638 | HEMA_IADH88 | P12669 influenza a |
| 639 | HEMA_IADH89 | P12670 influenza a |
| 640 | HEMA_IADH90 | P12671 influenza a |
| 641 | HEMA_IADH91 | P12672 influenza a |
| 642 | HEMA_IADH92 | P12673 influenza a |
| 643 | HEMA_IADH93 | P12674 influenza a |
| 644 | HEMA_IADH94 | P12675 influenza a |
| 645 | HEMA_IADH95 | P12676 influenza a |
| 646 | HEMA_IADH96 | P12677 influenza a |
| 647 | HEMA_IADH97 | P12678 influenza a |
| 648 | HEMA_IADH98 | P12679 influenza a |
| 649 | HEMA_IADH99 | P12680 influenza a |
| 650 | HEMA_IADH100 | P12681 influenza a |
| 651 | HEMA_IADH101 | P12682 influenza a |
| 652 | HEMA_IADH102 | P12683 influenza a |
| 653 | HEMA_IADH103 | P12684 influenza a |
| 654 | HEMA_IADH104 | P12685 influenza a |
| 655 | HEMA_IADH105 | P12686 influenza a |
| 656 | HEMA_IADH106 | P12687 influenza a |
| 657 | HEMA_IADH107 | P12688 influenza a |
| 658 | HEMA_IADH108 | P12689 influenza a |
| 659 | HEMA_IADH109 | P12690 influenza a |
| 660 | HEMA_IADH110 | P12691 influenza a |
| 661 | HEMA_IADH111 | P12692 influenza a |
| 662 | HEMA_IADH112 | P12693 influenza a |
| 663 | HEMA_IADH113 | P12694 influenza a |
| 664 | HEMA_IADH114 | P12695 influenza a |
| 665 | HEMA_IADH115 | P12696 influenza a |
| 666 | HEMA_IADH116 | P12697 influenza a |
| 667 | HEMA_IADH117 | P12698 influenza a |
| 668 | HEMA_IADH118 | P12699 influenza a |
| 669 | HEMA_IADH119 | P12700 influenza a |
| 670 | HEMA_IADH120 | P12701 influenza a |
| 671 | HEMA_IADH121 | P12702 influenza a |
| 672 | HEMA_IADH122 | P12703 influenza a |
| 673 | HEMA_IADH123 | P12704 influenza a |
| 674 | HEMA_IADH124 | P12705 influenza a |
| 675 | HEMA_IADH125 | P12706 influenza a |
| 676 | HEMA_IADH126 | P12707 influenza a |
| 677 | HEMA_IADH127 | P12708 influenza a |
| 678 | HEMA_IADH128 | P12709 influenza a |
| 679 | HEMA_IADH129 | P12710 influenza a |
| 680 | HEMA_IADH130 | P12711 influenza a |
| 681 | HEMA_IADH131 | P12712 influenza a |
| 682 | HEMA_IADH132 | P12713 influenza a |
| 683 | HEMA_IADH133 | P12714 influenza a |
| 684 | HEMA_IADH134 | P12715 influenza a |
| 685 | HEMA_IADH135 | P12716 influenza a |
| 686 | HEMA_IADH136 | P12717 influenza a |
| 687 | HEMA_IADH137 | P12718 influenza a |
| 688 | HEMA_IADH138 | P12719 influenza a |
| 689 | HEMA_IADH139 | P12720 influenza a |
| 690 | HEMA_IADH140 | P12721 influenza a |
| 691 | HEMA_IADH141 | P12722 influenza a |
| 692 | HEMA_IADH142 | P12723 influenza a |
| 693 | HEMA_IADH143 | P12724 influenza a |
| 694 | HEMA_IADH144 | P12725 influenza a |
| 695 | HEMA_IADH145 | P12726 influenza a |
| 696 | HEMA_IADH146 | P12727 influenza a |
| 697 | HEMA_IADH147 | P12728 influenza a |
| 698 | HEMA_IADH148 | P12729 influenza a |
| 699 | HEMA_IADH149 | P12730 influenza a |
| 700 | HEMA_IADH150 | P12731 influenza a |
| 701 | HEMA_IADH151 | P12732 influenza a |
| 702 | HEMA_IADH152 | P12733 influenza a |
| 703 | HEMA_IADH153 | P12734 influenza a |
| 704 | HEMA_IADH154 | P12735 influenza a |
| 705 | HEMA_IADH155 | P12736 influenza a |
| 706 | HEMA_IADH156 | P12737 influenza a |
| 707 | HEMA_IADH157 | P12738 influenza a |
| 708 | HEMA_IADH158 | P12739 influenza a |
| 709 | HEMA_IADH159 | P12740 influenza a |
| 710 | HEMA_IADH160 | P12741 influenza a |
| 711 | HEMA_IADH161 | P12742 influenza a |
| 712 | HEMA_IADH162 | P12743 influenza a |
| 713 | HEMA_IADH163 | P12744 influenza a |
| 714 | HEMA_IADH164 | P12745 influenza a |
| 715 | HEMA_IADH165 | P12746 influenza a |
| 716 | HEMA_IADH166 | P12747 influenza a |
| 717 | HEMA_IADH167 | P12748 influenza a |
| 718 | HEMA_IADH168 | P12749 influenza a |
| 719 | HEMA_IADH169 | P12750 influenza a |
| 720 | HEMA_IADH170 | P12751 influenza a |
| 721 | HEMA_IADH171 | P12752 influenza a |
| 722 | HEMA_IADH172 | P12753 influenza a |
| 723 | HEMA_IADH173 | P12754 influenza a |
| 724 | HEMA_IADH174 | P12755 influenza a |
| 725 | HEMA_IADH175 | P12756 influenza a |
| 726 | HEMA_IADH176 | P12757 influenza a |
| 727 | HEMA_IADH177 | P12758 influenza a |
| 728 | HEMA_IADH178 | P12759 influenza a |
| 729 | HEMA_IADH179 | P12760 influenza a |
| 730 | HEMA_IADH180 | P12761 influenza a |
| 731 | HEMA_IADH181 | P12762 influenza a |
| 732 | HEMA_IADH182 | P12763 influenza a |
| 733 | HEMA_IADH183 | P12764 influenza a |
| 734 | HEMA_IADH184 | P12765 influenza a |
| 735 | HEMA_IADH185 | P12766 influenza a |
| 736 | HEMA_IADH186 | P12767 influenza a |
| 737 | HEMA_IADH187 | P12768 influenza a |
| 738 | HEMA_IADH188 | P12769 influenza a |
| 739 | HEMA_IADH189 | P12770 influenza a |
| 740 | HEMA_IADH190 | P12771 influenza a |
| 741 | HEMA_IADH191 | P12772 influenza a |
| 742 | HEMA_IADH192 | P12773 influenza a |
| 743 | HEMA_IADH193 | P12774 influenza a |
| 744 | HEMA_IADH194 | P12775 influenza a |
| 745 | HEMA_IADH195 | P12776 influenza a |
| 746 | HEMA_IADH196 | P12777 influenza a |
| 747 | HEMA_IADH197 | P12778 influenza a |
| 748 | HEMA_IADH198 | P12779 influenza a |
| 749 | HEMA_IADH199 | P12780 influenza a |
| 750 | HEMA_IADH200 | P12781 influenza a |
| 751 | HEMA_IADH201 | P12782 influenza a |
| 752 | HEMA_IADH202 | P12783 influenza a |
| 753 | HEMA_IADH203 | P12784 influenza a |
| 754 | HEMA_IADH204 | P12785 influenza a |
| 755 | HEMA_IADH205 | P12786 influenza a |
| 756 | HEMA_IADH206 | P12787 influenza a |
| 757 | HEMA_IADH207 | P12788 influenza a |
| 758 | HEMA_IADH208 | P12789 influenza a |
| 759 | HEMA_IADH209 | P12790 influenza a |
| 760 | HEMA_IADH210 | P12791 influenza a |
| 761 | HEMA_IADH211 | P12792 influenza a |
| 762 | HEMA_IADH212 | P12793 influenza a |
| 763 | HEMA_IADH213 | P12794 influenza a |
| 764 | HEMA_IADH214 | P12795 influenza a |
| 765 | HEMA_IADH215 | P12796 influenza a |
| 766 | HEMA_IADH216 | P12797 influenza a |
| 767 | HEMA_IADH217 | P12798 influenza a |
| 768 | HEMA_IADH218 | P12799 influenza a |
| 769 | HEMA_IADH219 | P12800 influenza a |
| 770 | HEMA_IADH220 | P12801 influenza a |
| 771 | HEMA_IADH221 | P12802 influenza a |
| 772 | HEMA_IADH222 | P12803 influenza a |
| 773 | HEMA_IADH223 | P12804 influenza a |
| 774 | HEMA_IADH224 | P12805 influenza a |
| 775 | HEMA_IADH225 | P12806 influenza a |
| 776 | HEMA_IADH226 | P12807 influenza a |
| 777 | HEMA_IADH227 | P12808 influenza a |
| 778 | HEMA_IADH228 | P12809 influenza a |
| 779 | HEMA_IADH229 | P12810 influenza a |
| 780 | HEMA_IADH230 | P12811 influenza a |
| 781 | HEMA_IADH231 | P12812 influenza a |
| 782 | HEMA_IADH232 | P12813 influenza a |
| 783 | HEMA_IADH233 | P12814 influenza a |
| 784 | HEMA_IADH234 | P12815 influenza a |
| 785 | HEMA_IADH235 | P12816 influenza a |
| 786 | HEMA_IADH236 | P12817 influenza a |
| 787 | HEMA_IADH237 | P12818 influenza a |
| 788 | HEMA_IADH238 | P12819 influenza a |
| 789 | HEMA_IADH239 | P12820 influenza a |
| 790 | HEMA_IADH240 | P12821 influenza a |
| 791 | HEMA_IADH241 | P12822 influenza a |
| 792 | HEMA_IADH242 | P12823 influenza a |
| 793 | HEMA_IADH243 | P12824 influenza a |
| 794 | HEMA_IADH244 | P12825 influenza a |
| 795 | HEMA_IADH245 | P12826 influenza a |
| 796 | HEMA_IADH246 | P12827 influenza a |
| 797 | HEMA_IADH247 | P12828 influenza a |
| 798 | HEMA_IADH248 | P12829 influenza a |
| 799 | HEMA_IADH249 | P12830 influenza a |
| 800 | HEMA_IADH250 | P12831 influenza a |
| 801 | HEMA_IADH251 | P12832 influenza a |
| 802 | HEMA_IADH252 | P12833 influenza a |
| 803 | HEMA_IADH253 | P12834 influenza a |
| 804 | HEMA_IADH254 | P12835 influenza a |
| 805 | HEMA_IADH255 | P12836 influenza a |
| 806 | HEMA_IADH256 | P12837 influenza a |
| 807 | HEMA_IADH257 | P12838 influenza a |
| 808 | HEMA_IADH258 | P12839 influenza a |
| 809 | HEMA_IADH259 | P12840 influenza a |
| 810 | HEMA_IADH260 | P12841 influenza a |
| 811 | HEMA_IADH261 | P12842 influenza a |
| 812 | HEMA_IADH262 | P12843 influenza a |
| 813 | HEMA_IADH263 | P12844 influenza a |
| 814 | HEMA_IADH264 | P12845 influenza a |
| 815 | HEMA_IADH265 | P12846 influenza a |
| 816 | HEMA_IADH266 | P12847 influenza a |
| 817 | HEMA_IADH267 | P12848 influenza a |
| 818 | HEMA_IADH268 | P12849 influenza a |
| 819 | HEMA_IADH269 | P12850 influenza a |
| 820 | HEMA_IADH270 | P12851 influenza a |
| 821 | HEMA_IADH271 | P12852 influenza a |
| 822 | HEMA_IADH272 | P12853 influenza a |
| 823 | HEMA_IADH273 | P12854 influenza a |
| 824 | HEMA_IADH274 | P12855 influenza a |
| 825 | HEMA_IADH275 | P12856 influenza a |
| 826 | HEMA_IADH276 | P12857 influenza a |
| 827 | HEMA_IADH277 | P12858 influenza a |
| 828 | HEMA_IADH278 | P12859 influenza a |
| 829 | HEMA_IADH279 | P12860 influenza a |
| 830 | HEMA_IADH280 | P12861 influenza a |
| 831 | HEMA_IADH281 | P12862 influenza a |
| 832 | HEMA_IADH282 | P12863 influenza a |
| 833 | HEMA_IADH283 | P12864 influenza a |
| 834 | HEMA_IADH284 | P12865 influenza a |
| 835 | HEMA_IADH285 | P12866 influenza a |
| 836 | HEMA_IADH286 | P12867 influenza a |
| 837 | HEMA_IADH287 | P12868 influenza a |
| 838 | HEMA_IADH288 | P12869 influenza a |
| 839 | HEMA_IADH289 | P12870 influenza a |
| 840 | HEMA_IADH290 | P12871 influenza a |
| 841 | HEMA_IADH291 | P12872 influenza a |
| 842 | HEMA_IADH292 | P12873 influenza a |
| 843 | HEMA_IADH293 | P12874 influenza a |
| 844 | HEMA_IADH294 | P12875 influenza a |
| 845 | HEMA_IADH295 | P12876 influenza a |
| 846 | HEMA_IADH296 | P12877 influenza a |
| 847 | HEMA_IADH297 | P12878 influenza a |
| 848 | HEMA_IADH298 | P12879 influenza a |
| 849 | HEMA_IADH299 | P12880 influenza a |
| 850 | HEMA_IADH300 | P12881 influenza a |
| 851 | HEMA_IADH301 | P12882 influenza a |
| 852 | HEMA_IADH302 | P12883 influenza a |
| 853 | HEMA_IADH303 | P12884 influenza a |
| 854 | HEMA_IADH304 | P12885 influenza a |
| 855 | HEMA_IADH305 | P12886 influenza a |
| 856 | HEMA_IADH306 | P12887 influenza a |
| 857 | HEMA_IADH307 | P12888 influenza a |
| 858 | HEMA_IADH308 | P12889 influenza a |
| 859 | HEMA_IADH309 | P12890 influenza a |
| 860 | HEMA_IADH310 | P12891 influenza a |
| 861 | HEMA_IADH311 | P12892 influenza a |
| 862 | HEMA_IADH312 | P12893 influenza a |
| 863 | HEMA_IADH313 | P12894 influenza a |
| 864 | HEMA_IADH314 | P12895 influenza a |
| 865 | HEMA_IADH315 | P12896 influenza a |
| 866 | HEMA_IADH316 | P12897 influenza a |
| 867 | HEMA_IADH317 | P12898 influenza a |
| 868 | HEMA_IADH318 | P12899 influenza a |
| 869 | HEMA_IADH319 | P12900 influenza a |
| 870 | HEMA_IADH320 | P12901 influenza a |
| 871 | HEMA_IADH321 | P12902 influenza a |
| 872 | HEMA_IADH322 | P12903 influenza a |
| 873 | HEMA_IADH323 | P12904 influenza a |
| 874 | HEMA_IADH324 | P12905 influenza a |
| 875 | HEMA_IADH325 | P12906 influenza a |
| 876 | HEMA_IADH326 | P12907 influenza a |
| 877 | HEMA_IADH327 | P12908 influenza a |
| 878 | HEMA_IADH328 | P12909 influenza a |
| 879 | HEMA_IADH329 | P12910 influenza a |
| 880 | HEMA_IADH330 | P12911 influenza a |
| 881 | HEMA_IADH331 | P12912 influenza a |
| 882 | HEMA_IADH332 | P12913 influenza a |
| 883 | HEMA_IADH333 | P12914 influenza a |
| 884 | HEMA_IADH334 | P12915 influenza a |
| 885 | HEMA_IADH335 | P12916 influenza a |
| 886 | HEMA_IADH336 | P12917 influenza a |
| 887 | HEMA_IADH337 | P12918 influenza a |
| 888 | HEMA_IADH338 | P12919 influenza a |
| 889 | HEMA_IADH339 | P12920 influenza a |
| 890 | HEMA_IADH340 | P12921 influenza a |
| 891 | HEMA_IADH341 | P12922 influenza a |
| 892 | HEMA_IADH342 | P12923 influenza a |
| 893 | HEMA_IADH343 | P12924 influenza a |
| 894 | HEMA_IADH344 | P12925 influenza a |
| 895 | HEMA_IADH345 | P12926 influenza a |
| 896 | HEMA_IADH346 | P12927 influenza a |

| | | | | | | | | | | | | | | | |
|-----|---|-----|-----|---|-------------|---------|--------------|-----|---|-----|-----|---|------------|--------|--------------|
| 107 | 6 | 2.5 | 206 | 1 | SOD2_PLEBO | P50059 | plectonema | 180 | 6 | 2.5 | 350 | 1 | YN14_YEAST | P53939 | saccharomyc |
| 108 | 6 | 2.5 | 207 | 1 | JAG_EACHD | Q9rcas | bacillus ha | 181 | 6 | 2.5 | 356 | 1 | KPRA_HUMAN | Q14558 | homo sapien |
| 109 | 6 | 2.5 | 209 | 1 | NGM_PART2 | P15600 | paramacium | 182 | 6 | 2.5 | 356 | 1 | KPRA_MOUSE | Q9dum1 | mus musculus |
| 110 | 6 | 2.5 | 210 | 1 | GUSTI_CABEL | Q09607 | caenorhabdi | 183 | 6 | 2.5 | 356 | 1 | KPRA_RAT | Q63468 | rattus norv |
| 111 | 6 | 2.5 | 213 | 1 | RS3_OCBH | P59182 | oceanobacil | 184 | 6 | 2.5 | 361 | 1 | PAX1_HUMAN | P15863 | homo sapien |
| 112 | 6 | 2.5 | 214 | 1 | SODF_ARPE | Q9v8h8 | aeropyrum p | 185 | 6 | 2.5 | 369 | 1 | DNAJ_HELPJ | Q9ziq2 | helicobacte |
| 113 | 6 | 2.5 | 217 | 1 | FLA1_METJA | Q58301 | methanococc | 186 | 6 | 2.5 | 369 | 1 | DNAJ_HELPY | Q25850 | helicobacte |
| 114 | 6 | 2.5 | 217 | 1 | FLA2_METJA | Q58302 | methanococc | 187 | 6 | 2.5 | 369 | 1 | KPRB_HUMAN | Q60256 | homo sapien |
| 115 | 6 | 2.5 | 217 | 1 | RS3_EAST | P23309 | bacillus st | 188 | 6 | 2.5 | 369 | 1 | KPRB_MOUSE | Q8r374 | mus musculus |
| 116 | 6 | 2.5 | 217 | 1 | RS3_BACSU | P21465 | bacillus su | 189 | 6 | 2.5 | 369 | 1 | KPRB_RAT | Q08618 | rattus norv |
| 117 | 6 | 2.5 | 217 | 1 | RS3_LACLU | Q9cdw8 | lactococcus | 190 | 6 | 2.5 | 371 | 1 | CYB_LATCO | Q9mj14 | laticauda c |
| 118 | 6 | 2.5 | 217 | 1 | RS3_STAM | Q99827 | staphylococ | 191 | 6 | 2.5 | 371 | 1 | DUT_HSV11 | P10234 | herpes simp |
| 119 | 6 | 2.5 | 217 | 1 | RS3_STAEF | Q8crg6 | staphylococ | 192 | 6 | 2.5 | 373 | 1 | CD62_METAC | Q8tui2 | methanoea |
| 120 | 6 | 2.5 | 217 | 1 | RS3_STRMU | P59186 | streptococc | 193 | 6 | 2.5 | 375 | 1 | DNAJ_UREPA | Q9pc82 | ureaplasma |
| 121 | 6 | 2.5 | 217 | 1 | RS3_STRPN | Q9w437 | streptococc | 194 | 6 | 2.5 | 375 | 1 | SR55_DROME | P26686 | drosophila |
| 122 | 6 | 2.5 | 217 | 1 | RS3_STRPY | Q9aiw8 | streptococc | 195 | 6 | 2.5 | 379 | 1 | DNAJ_LSGPN | P50025 | legionella |
| 123 | 6 | 2.5 | 218 | 1 | RS3_LISMO | Q92713 | listeria mo | 196 | 6 | 2.5 | 388 | 1 | OVAY_CHICK | P01014 | gallus gall |
| 124 | 6 | 2.5 | 219 | 1 | RS3_BACAA | Q81vsa4 | bacillus an | 197 | 6 | 2.5 | 388 | 1 | YC09_KLEPN | Q48455 | klebsiella |
| 125 | 6 | 2.5 | 219 | 1 | RS3_BACCR | Q81j36 | bacillus ce | 198 | 6 | 2.5 | 390 | 1 | TRPB_METTM | P26921 | methanobact |
| 126 | 6 | 2.5 | 219 | 1 | RS3_BACHD | Q9z9k8 | bacillus ha | 199 | 6 | 2.5 | 392 | 1 | TRB1_METTH | Q27696 | methanobact |
| 127 | 6 | 2.5 | 222 | 1 | SODM_HORSE | Q9x841 | equus cabal | 200 | 6 | 2.5 | 398 | 1 | Y095_MYCGE | P47341 | mycoplasma |
| 128 | 6 | 2.5 | 223 | 1 | PGC2_HUMAN | O15173 | homo sapien | | | | | | | | |
| 129 | 6 | 2.5 | 223 | 1 | THTP_RAT | Q8cgv7 | rattus norv | | | | | | | | |
| 130 | 6 | 2.5 | 235 | 1 | LIPB_MYLEE | C32961 | mycobacteri | | | | | | | | |
| 131 | 6 | 2.5 | 248 | 1 | SOD1_PLEBO | P50058 | plectonema | | | | | | | | |
| 132 | 6 | 2.5 | 250 | 1 | TRYP_PLEPL | P35034 | pleuronecte | | | | | | | | |
| 133 | 6 | 2.5 | 260 | 1 | GRAA_MOUSE | P11032 | mus musculus | | | | | | | | |
| 134 | 6 | 2.5 | 262 | 1 | GRAA_HUMAN | P12544 | homo sapien | | | | | | | | |
| 135 | 6 | 2.5 | 267 | 1 | ELNE_HUMAN | P08246 | homo sapien | | | | | | | | |
| 136 | 6 | 2.5 | 268 | 1 | ZUPT_OCEIH | Q8enq1 | oceanobacil | | | | | | | | |
| 137 | 6 | 2.5 | 273 | 1 | YDGB_ECOLI | P76176 | escherichia | | | | | | | | |
| 138 | 6 | 2.5 | 281 | 1 | UPK_STRPN | Q97sc8 | streptococc | | | | | | | | |
| 139 | 6 | 2.5 | 282 | 1 | AQP6_HUMAN | P11520 | homo sapien | | | | | | | | |
| 140 | 6 | 2.5 | 287 | 1 | TRUB_FUSNN | Q8r5x8 | fusobacteri | | | | | | | | |
| 141 | 6 | 2.5 | 291 | 1 | ZUPT_CAMJE | Q9pin2 | campylobact | | | | | | | | |
| 142 | 6 | 2.5 | 292 | 1 | EFTS_XANAC | Q8pmk6 | xanthomonas | | | | | | | | |
| 143 | 6 | 2.5 | 292 | 1 | EFTS_XANCP | Q8pav3 | xanthomonas | | | | | | | | |
| 144 | 6 | 2.5 | 292 | 1 | EFTS_XYLFA | Q9pad9 | xyella fas | | | | | | | | |
| 145 | 6 | 2.5 | 292 | 1 | EFTS_XYLFT | Q87a70 | xyella fas | | | | | | | | |
| 146 | 6 | 2.5 | 298 | 1 | VGLR_HRSV5 | P27024 | human respi | | | | | | | | |
| 147 | 6 | 2.5 | 299 | 1 | GRF4_HUMAN | Q9g227 | homo sapien | | | | | | | | |
| 148 | 6 | 2.5 | 301 | 1 | CAPR_VERPE | P26950 | versinia pe | | | | | | | | |
| 149 | 6 | 2.5 | 302 | 1 | Y548_STAEP | Q8cte3 | staphylococ | | | | | | | | |
| 150 | 6 | 2.5 | 303 | 1 | Y367_RICPR | Q9z8g2 | ricketsia | | | | | | | | |
| 151 | 6 | 2.5 | 308 | 1 | HEM2_RHILO | Q98e17 | rhizobium l | | | | | | | | |
| 152 | 6 | 2.5 | 311 | 1 | FMT_BRAJA | Q89dp0 | bradyrhizob | | | | | | | | |
| 153 | 6 | 2.5 | 313 | 1 | M20M_BOVIN | P22292 | bos taurus | | | | | | | | |
| 154 | 6 | 2.5 | 313 | 1 | M20M_HUMAN | Q02978 | homo sapien | | | | | | | | |
| 155 | 6 | 2.5 | 313 | 1 | M20M_MOUSE | Q9cr62 | mus musculus | | | | | | | | |
| 156 | 6 | 2.5 | 313 | 1 | M20M_RAT | P97700 | rattus norv | | | | | | | | |
| 157 | 6 | 2.5 | 313 | 1 | MRWV_PSESM | Q87wx7 | pseudomonas | | | | | | | | |
| 158 | 6 | 2.5 | 313 | 1 | MRWV_PSEPK | Q88n84 | pseudomonas | | | | | | | | |
| 159 | 6 | 2.5 | 316 | 1 | LDH_STAEP | Q8cmz0 | staphylococ | | | | | | | | |
| 160 | 6 | 2.5 | 325 | 1 | YJH3_MYCTU | Q10552 | mycobacteri | | | | | | | | |
| 161 | 6 | 2.5 | 326 | 1 | YJH5_ECOLI | P39370 | escherichia | | | | | | | | |
| 162 | 6 | 2.5 | 328 | 1 | P2Y6_HUMAN | Q15077 | homo sapien | | | | | | | | |
| 163 | 6 | 2.5 | 329 | 1 | HOLB_BACSU | P37540 | bacillus su | | | | | | | | |
| 164 | 6 | 2.5 | 331 | 1 | PME_ASPAC | Q12535 | aspergillus | | | | | | | | |
| 165 | 6 | 2.5 | 332 | 1 | OPPF_HAEIN | P45051 | haemophilus | | | | | | | | |
| 166 | 6 | 2.5 | 334 | 1 | OPPF_ECOLI | P77737 | escherichia | | | | | | | | |
| 167 | 6 | 2.5 | 334 | 1 | OPPF_SALTY | P08007 | salmonella | | | | | | | | |
| 168 | 6 | 2.5 | 334 | 1 | TRPD_STRPN | Q97p29 | streptococc | | | | | | | | |
| 169 | 6 | 2.5 | 336 | 1 | CMST_CRIGR | Q08520 | cricetulus | | | | | | | | |
| 170 | 6 | 2.5 | 336 | 1 | CMST_MOUSE | Q61420 | mus musculus | | | | | | | | |
| 171 | 6 | 2.5 | 337 | 1 | CMST_HUMAN | P78382 | homo sapien | | | | | | | | |
| 172 | 6 | 2.5 | 339 | 1 | ANM1_SCHPO | Q9urx7 | schizosacch | | | | | | | | |
| 173 | 6 | 2.5 | 344 | 1 | ETFA_YEAST | Q12480 | saccharomyc | | | | | | | | |
| 174 | 6 | 2.5 | 346 | 1 | OMPA_ECOLI | P02934 | escherichia | | | | | | | | |
| 175 | 6 | 2.5 | 346 | 1 | YE97_METJA | Q58892 | methanococc | | | | | | | | |
| 176 | 6 | 2.5 | 348 | 1 | YGD9_YEAST | P53183 | saccharomyc | | | | | | | | |
| 177 | 6 | 2.5 | 349 | 1 | YCXF_PORPU | P51277 | porphyra pu | | | | | | | | |
| 178 | 6 | 2.5 | 350 | 1 | OMPA_ENTAE | P09146 | enterobacte | | | | | | | | |
| 179 | 6 | 2.5 | 350 | 1 | OMPA_SALTY | P02936 | salmonella | | | | | | | | |

ALIGNMENTS

RESULT 1

CA34_HUMAN STANDARD; PRT; 1670 AA.
 ID CA34_HUMAN
 AC Q01955; Q9BQT2;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
 GN COL4A3
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney; PubMed=8083201;
 EX MEDLINE=94364994; PubMed=8083201;
 RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.
 RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
 RT human tissues.";
 RL J. Biol. Chem. 269:23013-23017(1994).
 RN [2]
 RP REVISIONS
 RA Leinonen A.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-1167;
 RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
 RP CYS-1661; AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
 RP PRO-574; GLU-1269 AND PRO-1474.
 RX MEDLINE=21064696; PubMed=1134255;
 RA Heidt L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
 RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in
 RT autosomal Alport syndrome.";
 RL J. Am. Soc. Nephrol. 12:97-106(2001).
 RN [4]
 RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=93015826; PubMed=1400291;
 RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
 RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
 RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
 RT antigenic region at the triple helix/NC1 domain junction.";
 RL J. Biol. Chem. 267:19780-19784(1992).
 RN [5]
 RP SEQUENCE OF 1453-1670 FROM N.A.

RX MEDLINE=91353570; PubMed=1882840;
RA Morrison K.E., Mariyama M., Yang-Peng T.L., Readers S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147678; PubMed=1737849;
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601(1992).
RN [7]
RP SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=Kidney;
RA Ding J.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NCI domain of the human alpha 3 (IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [9]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Nimomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3 (IV) and
RT alpha4 (IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [10]
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [11]
RP VARIANT PRO-1474.
RX MEDLINE=95078927; PubMed=7987301;
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Readers S.T., Smets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT recessive Alport syndrome.";
RL Hum. Mol. Genet. 3:1269-1273(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist. Isoforms differ in
CC the C-terminal part of the NCI domain;
CC Name=1;
CC IsoId=Q01955-1; Sequences=Displayed;
CC Name=2; Synonyms=V;
CC IsoId=Q01955-2; Sequences=VSP_001170;
CC Name=3; Synonyms=L5;
CC IsoId=Q01955-3; Sequences=VSP_001171;
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,

```
DR EMBL; AJ288528; CAC36101.1; JOINED.
DR EMBL; AJ288529; CAC36101.1; JOINED.
DR EMBL; AJ288530; CAC36101.1; JOINED.
DR EMBL; AJ288531; CAC36101.1; JOINED.
DR EMBL; AJ288532; CAC36101.1; JOINED.
DR EMBL; AJ288533; CAC36101.1; JOINED.
DR EMBL; AJ288534; CAC36101.1; JOINED.
DR EMBL; AJ288535; CAC36101.1; JOINED.
DR EMBL; AJ288536; CAC36101.1; JOINED.

Query Match 100.0%; Score 244; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 1.5e-242;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKXKRGDSGSPATWTRGFVTRHSQTTAIFSCPEGTVPLYSFGSFVQGNQRAHQD 60
DB 1427 GLKXKRGDSGSPATWTRGFVTRHSQTTAIFSCPEGTVPLYSFGSFVQGNQRAHQD 1486

QY 61 LGTLGSLQRFMTPELFCNVNDVNCNFASTRNDYSYWLSTPALMPNMNAPITGRALEPYIS 120
DB 1487 LGTLGSLQRFMTPELFCNVNDVNCNFASTRNDYSYWLSTPALMPNMNAPITGRALEPYIS 1546

QY 121 RCTVCEGPAIAVAHSQTTDIPCPHGWISLWKGSFIMFTSAGSECTGOALASPGSCLE 180
DB 1547 RCTVCEGPAIAVAHSQTTDIPCPHGWISLWKGSFIMFTSAGSECTGOALASPGSCLE 1606

QY 181 EFRASPFLECHRGTCNNYSNSYSFWLASINPERMFRKPIPTVXAGELEKTIISRCQVCM 240
DB 1607 EFRASPFLECHRGTCNNYSNSYSFWLASINPERMFRKPIPTVXAGELEKTIISRCQVCM 1666

QY 241 KKRH 244
DB 1667 KKRH 1670

RESULT 2
CA34_BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3 (IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Readers S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a CDNA
encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
alpha 4, of type IV collagen."
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3 (IV) chain
of collagen IV."
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]

SEQUENCE OF 227-244.
MEDLINE=87222419; PubMed=2438283;
RA Rutkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
membrane collagen."
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1 (IV) -
alpha 6 (IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -1- PM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63139; AAA62708.1; -.
CC PIR; A39024; A39024.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagen4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 4.
CC ProDom; PD003923; ProcollagenC4; 1.
CC SMART; SM00111; C4; 2.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
FT NON_TER 1 1 Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT SITE 239 471 NONHELICAL REGION (NC1).
FT MOD_RES 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 465 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match 16.0%; Score 39; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 4e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAVAHSQTTDIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAVAHSQTTDIPPCP 372

RESULT 3
CA54_CANFA STANDARD; PRT; 754 AA.
ID -CA54_CANFA
```

Q28247;
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Collagen alpha 5(IV) chain (Fragment).
 COL4A5.
 Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Samoyed; TISSUE=Kidney;
 RX MEDLINE=94224868; PubMed=8171024;
 RA "Canine X chromosome-linked hereditary nephritis: a genetic model for
 human X-linked hereditary nephritis resulting from a single base
 mutation in the gene encoding the alpha 5 chain of collagen type
 IV.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'
 CC meshwork together with laminins, proteoglycans and entactin/
 CC nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
 CC alpha 6(IV), each of which can form a triple helix structure with
 CC 2 other chains to generate type IV collagen network.
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-
 CC X-Y repeats in the long central triple-helical domain (which may
 CC cause flexibility in the triple helix), and a short N-terminal
 CC triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 CC are involved in inter- and intramolecular disulfide bonding. 12 of
 CC these, located in the NC1 domain, are conserved in all known type
 CC IV collagens.
 CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
 CC canine X-linked hereditary nephritis (HN), a disease similar to
 CC that in humans (also referred to as Alport syndrome) characterized
 CC by progressive renal failure and neurosensory deafness.
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; U07888; AAB60258.1; -;
 DR PIR; A55267; A55267.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 8.
 DR ProDom; PD000007; Clg_helix; 1.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 FT NON_TER 1 1
 FT DOMAIN <1 530 TRIPLE-HELICAL REGION.
 FT DOMAIN 531 >754 NONHELICAL REGION (NC1).
 FT DISULFID 552 643 OR 640 (BY SIMILARITY).
 FT DISULFID 585 640 OR 643 (BY SIMILARITY).
 FT DISULFID 597 603 BY SIMILARITY.
 FT DISULFID 662 ? OR 754 (BY SIMILARITY).
 FT DISULFID 696 754 BY SIMILARITY.

FT DISULFID 708 714 BY SIMILARITY.
 FT NON_TER 754 754
 SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;
 Query Match 7.0%; Score 17; DB 1; Length 754;
 Best Local Similarity 100.0%; Pred. No. 2.6e-09; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 0;
 Qy 84 VCNFASRNDYSYWLSTP 100
 |||||
 Db 602 VCNFASRNDYSYWLSTP 618
 RESULT 4
 CA14 HUMAN
 ID CA14 HUMAN STANDARD; PRT; 1669 AA.
 AC P02462;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340433; PubMed=2701944;
 RA Soiminen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RT type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [2]
 RP SEQUENCE OF 46-1257 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=8803584; PubMed=3691802;
 RA Soiminen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RT "Complete primary structure of the alpha 1-chain of human basement
 RT membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [3]
 RP SEQUENCE OF 1-943 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88029471; PubMed=3311751;
 RA Brazel D., Oberhauser I., Dieringer H., Babel W., Glanville R.W.,
 RT Deutzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 RT basement membrane collagen (type IV) reveals 21 non-triplet
 RT interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN [4]
 RP SEQUENCE OF 28-243
 RX MEDLINE=8604708; PubMed=4043082;
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking
 RT region (7S domain) of the alpha 1 (IV) chain of human basement
 RT membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN [5]
 RP SEQUENCE OF 534-1447.
 RX MEDLINE=85003629; PubMed=6434307;
 RA Babel W., Glanville R.W.;
 RT "Structure of human-basement-membrane (type IV) collagen. Complete
 RT amino-acid sequence of a 914-residue-long pepsin fragment from the
 RT alpha 1(IV) chain.";
 RL Eur. J. Biochem. 143:545-556(1984).
 RN [6]
 RP SEQUENCE OF 1256-1669 FROM N.A.
 RX MEDLINE=85207819; PubMed=2581969;
 RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV

RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2982422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Lysines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26576; AAA53098.1; JOINED.
DR EMBL; J04217; AAA53098.1; JOINED.
DR EMBL; M26550; AAA53098.1; JOINED.
DR EMBL; M26540; AAA53098.1; JOINED.
DR EMBL; M26542; AAA53098.1; JOINED.
DR EMBL; M26543; AAA53098.1; JOINED.
DR EMBL; M26544; AAA53098.1; JOINED.
DR EMBL; M26545; AAA53098.1; JOINED.
DR EMBL; M26546; AAA53098.1; JOINED.
DR EMBL; M26547; AAA53098.1; JOINED.
DR EMBL; M26537; AAA53098.1; JOINED.
DR EMBL; M26538; AAA53098.1; JOINED.
DR EMBL; M26548; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.

DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.
DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA86698.1; -
DR EMBL; X05561; CAA23075.1; -
DR EMBL; M10940; AAA52006.1; -
DR EMBL; M11315; AAA52042.1; -
DR PR; SI6876; CGHU4B.
DR Genew; HGNC:2202; COL4A1.
DR MM; 120130; -
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; Clg helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1662.
FT DISULFID 1570 1665 OR 1662.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1622
FT CONFLICT 237 238 SG -> KE (IN REF. 4).
FT CONFLICT 241 241 Q -> K (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SQ SEQUENCE 1669 AA; 160611 MW; 3BEBAGDFFB9B8A84 CRC64;

Query Match 7.0%; Score 17; DB 1; Length 1669;
Best Local Similarity 100.0%; Pred. No. 5.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1622 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 7.0%; Score 17; DB 1; Length 1669;
Best Local Similarity 100.0%; Pred. No. 5,1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
Db 1510 VCNFASRNDYSYWLSTP 1526

RESULT 6
ID#_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
AT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5 (IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
[2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
[3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjola-Erila E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5 (IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
[4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
[5]

```

```

RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeslyhyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
[6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjola-Erila E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5 (IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
[7]
RP SEQUENCE OF 1-20 FROM N.A.
RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RA Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NCI domain.";
RL Kidney Int. 44:1318-1321(1993).
[9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
[10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5 (IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
[11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: Characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
[12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
[13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
[14]

```

RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;
RP ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=9532976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
RN [15]
RP VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
RN [16]
RP VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
RN [17]
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=97094179; PubMed=8940267;
RA Knebelmann B., Brallat C., Forestier L., Arrondel C., Jacassier D.,
RA Glatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
RN [18]
RP VARIANT AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobrynya V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of $\alpha 5$ (IV) collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
RN [19]
RP VARIANTS AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q";
RL Hum. Genet. 99:681-684(1997).
RN [20]
RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
RN [21]
RP VARIANTS AS VAL-420; 456-PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP 802-GLY-PRO-807 DEL; ARG-889; CYS-941; SER-1030; SER-1066; ASP-1143;
RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
RX MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
RN [22]
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;

RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
RT sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
RN [23]
RP VARIANT AS ARG-822.
Query Match 7.0%; Score 17; DB 1; Length 1685;
Best Local Similarity 100.0%; Pred. No. 5.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
DB 1526 VCNFASRNDYSYWLSTP 1542
CA14_CABE1_CABE1 STANDARD; PRT; 1758 AA.
AC P17139;
DC 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CLB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
RT collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lighthouse J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RX STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -1- FUNCTION: Collagen type IV is specific for basement membranes.
CC -1- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.

Type IV collagen forms a mesh-like network linked through intermolecular interactions between 7S domains and between NC1 domains.

-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

-!- PPM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

-!- PPM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

-!- DISEASE: Mutations in this gene cause temperature-sensitive lethality during late embryogenesis.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; X56979; CAA40299.1; -;
EMBL; Z27078; CAA81584.3; -;
EMBL; J05067; AAB59179.1; -;
PIR; S40991; S40991;
WormPep; K04H4.1; CE32462.
InterPro; IPR008161; C1g.Helix.
InterPro; IPR008160; Collagen.
InterPro; IPR001442; Procollagen4_C.
Pfam; PF01413; C4; 2.
Pfam; PF01391; Collagen; 27.
ProDom; PD000007; C1g.Helix; 11.
ProDom; PD003923; Procollagen4; 1.
SMART; SM00111; C4; 2.
Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Signal.
POTENTIAL.
SIGNAL 1 20
PROPEP 21 2194 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
CHAIN 2195 1758 COLLAGEN ALPHA 1(IV) CHAIN.
DOMAIN 195 1529 TRIPLE-HELICAL REGION.
DOMAIN 1530 1758 NON-HELICAL REGION (NC1).
DISULFID 1549 1640 OR 1637 (BY SIMILARITY).
DISULFID 1582 1637 OR 1640 (BY SIMILARITY).
DISULFID 1594 1600 BY SIMILARITY.
DISULFID 1659 1754 OR 1751 (BY SIMILARITY).
DISULFID 1693 1751 OR 1754 (BY SIMILARITY).
DISULFID 1705 1711 BY SIMILARITY.
VARIANT 402 402 G -> E (IN MUTANT G34).
VARIANT 408 408 G -> E (IN MUTANT G23/HG70).
CONFLICT 302 391 LDNGKRGKGVGNGYGEKSGQGLGGTGPYPTKGGAGE
PGYPRGFGSDGCGPEGLGEGTGEAGPHCAQGFQGVQSGK
GLPFGDGL -> AGQVSIQPNKKLFLFCRVNVERTEQSE
IMKRPDKNKDELQDTQLREGLNQNQTDQDSKRTV
DRKDLKELKRLVLDHMLKDSITFEKAKCQDMVVS (IN
REF. 2).
CONFLICT 581 581 G -> R (IN REF. 2).
CONFLICT 768 768 P -> R (IN REF. 2).
CONFLICT 830 830 D -> V (IN REF. 2).
CONFLICT 1514 1514 P -> Q (IN REF. 4).
CONFLICT 1722 1722 P -> L (IN REF. 2).
SEQUENCE 1758 AA; 170857 MW; 7083D9AF63E05D45 CRC64;
Query Match 4.5%; Score 11; DB 1; Length 1758;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 174 SPGSCLEEFRA 184
Db 1689 SPGSCLEEFRA 1699

RESULT 8
CA24 HUMAN
ID CA24 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
collagen and comparison with the alpha 1(IV) chain."
RL J. Biol. Chem. 263:19488-19493(1988).
RN [2]
RP SEQUENCE OF 1-1042 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88151998; PubMed=3345760;
RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;
RT "Human basement membrane collagen (type IV). The amino acid sequence
of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
chain reveals deletions in the alpha 1(IV) chain."
RL Eur. J. Biochem. 172:35-42(1988).
RN [3]
RP SEQUENCE OF 1254-1712 FROM N.A.
RX MEDLINE=87219158; PubMed=3582677;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2
chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
(IV) but not with the collagenous domain or 3'-untranslated region."
RL FEBS Lett. 216:281-286(1987).
RN [4]
RP SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
alpha 2 (IV) chains map to the distal long arm of chromosome 13."
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
RN [5]
RP SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-
specific expression of alpha 1(IV) and alpha 2(IV) collagen genes."
RL J. Biol. Chem. 262:9231-9238(1987).
RN [6]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Sohininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
collagen are divergently encoded on opposite DNA strands and have an
overlapping promoter region."
RL J. Biol. Chem. 263:17217-17220(1988).
RN [7]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89030632; PubMed=2846280;
RA Poeschl E., Pollner R., Kuehn K.;
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
basement membrane collagen type IV are arranged head-to-head and
separated by a bidirectional promoter of unique structure."
RL EMBO J. 7:2687-2695(1988).
RN [8]
RP SEQUENCE OF 1-33 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=93305049; PubMed=8317999;

RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
RT "Identification of a novel sequence element in the common promoter
RT region of human collagen type IV genes, involved in the regulation of
RT divergent transcription.";
RL Biochem. J. 292:687-695(1993).
RN [9]
RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05562; CAA29076.1; -;
DR EMBL; X05610; CAA29098.1; -;
DR EMBL; J02760; AAA58422.1; -;
DR EMBL; M36963; AAA53059.1; -;
DR EMBL; X12784; CAA31275.1; -;
DR EMBL; J04217; AAA53097.1; -;
DR PIR; A32024; CGH2B.
DR Genew; HGNC:2203; COL4A2.
DR MIM; 120030; -;
DR GO; GO:0005587; C:collagen type IV; TAS.
DR GO; GO:0005201; P:extracellular matrix structural constituent; TAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. . .; NAS.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 7.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25
FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.
FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
FT DISULFID 1337 1590 OR 1593 (BY SIMILARITY).
FT DISULFID 1349 1555 BY SIMILARITY.
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).

FT DISULFID 1658 1665 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
RT R -> P (IN REF. 2).
FT CONFLICT 471 471 A -> G (IN REF. 2).
FT CONFLICT 683 683 M -> I (IN REF. 5).
FT CONFLICT 1575 1575 G -> H (IN REF. 9).
FT CONFLICT 1663 1663 H -> G (IN REF. 9).
FT CONFLICT 1701 1701 H -> G (IN REF. 9).
SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
Query Match 4.1%; Score 10; DB 1; Length 1712;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 128 PAIAIAVHSQ 137
Db 1596 PAIAIAVHSQ 1605
RESULT 9
CA24 MOUSE
ID CA24 MOUSE STANDARD; PRT; 1707 AA.
AC P08122; Q61375;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; PubMed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
RA Pihlajaniemi T., Kurkinen M.;
RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen.";
RL J. Biol. Chem. 264:6318-6324(1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
RA Deutzmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RT chain and its comparison with the alpha 1(IV) chain.";
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1(IV).";
RL FEBS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;

"Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";

FEBS Lett. 206:29-32(1986).

[7]

SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.

RX MEDLINE=85296379; PubMed=3839908;

RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;

RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";

RL Nature 317:177-179(1985).

[8]

SEQUENCE OF 1-60 FROM N.A.

RX MEDLINE=89071759; PubMed=3200851;

RA Burello P.D., Martin G.R., Yamada Y.;

RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";

RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).

CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) - alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M23334; AAA51626.1; -

DR EMBL; M23333; AAA51626.1; JOINED.

DR EMBL; J04695; AAA50293.1; -

DR EMBL; J04448; AAA37438.1; -

DR EMBL; X04647; CRA28308.1; -

DR EMBL; M15833; AAA37341.1; -

DR EMBL; X04410; CRA27998.1; -

DR EMBL; X02896; CRA26655.1; -

DR EMBL; X02897; CAB51614.1; -

DR EMBL; X02898; CRA26657.1; -

DR EMBL; X02899; CRA26658.1; -

DR PIR; A33526; A33526.

DR MGD; MGI:38455; Col4a2.

DR GO; GO:005604; C:basement membrane; IDA.

DR InterPro; IPR008161; C1g helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR001442; Procollagen4_C.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 21.

DR ProDom; PD000007; C1g helix; 7.

DR ProDom; PD003923; ProCollagenC4; 1.

DR SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Signal.

FT SIGNAL 1 25

FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 184 1707 COLLAGEN ALPHA 2(IV) CHAIN.

FT DOMAIN 184 1479 TRIPLE-HELICAL REGION.

FT DOMAIN 1480 1707 NONHELICAL REGION (NC1).

FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).

FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).

FT DISULFID 1544 1550 BY SIMILARITY.

FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).

FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).

FT DISULFID 1653 1660 BY SIMILARITY.

FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1370 1270 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 1051 1051 P -> R (IN REF. 6).

FT CONFLICT 1097 1097 S -> G (IN REF. 7).

FT CONFLICT 1171 1171 G -> S (IN REF. 6).

FT CONFLICT 1179 1179 P -> R (IN REF. 6).

FT CONFLICT 1241 1241 Q -> E (IN REF. 6).

FT CONFLICT 1328 1328 P -> A (IN REF. 6).

FT CONFLICT 1573 1573 V -> L (IN REF. 4).

FT CONFLICT 1623 1623 Y -> H (IN REF. 4).

SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;

Query Match 3.7%; Score 9; DB 1; Length 1707;

Best Local Similarity 100.0%; Pred. No. 0.88;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISRCQVCMK 241

Db 1697 ISRCQVCMK 1705

RESULT 10

SODM AGABI STANDARD; PRT; 200 AA.

AC Q2P4T6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Superoxide dismutase [Mn], mitochondrial precursor (RC 1.15.1.1).

GN SOD.

OS Agaricus bisporus (Common mushroom).

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Agaricales; Agaricaceae; Agaricus.

OX NCBI_TaxID=5341;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Horst U3;

RA Eastwood D.C., Bains N.K., Henderson J., Burton K.S.;

RL "Oxidative stress in the harvested mushroom, *Agaricus bisporus*.";

RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Destroys radicals which are normally produced within the cells and which are toxic to biological systems.

CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).

CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.

CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase family.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ404469; CAB94731.1; -

DR HSSP; P04179; 1VAR.

DR InterPro; IPR001189; SODismutase.

DR Pfam; PF00881; sodfe; 1.

DR Pfam; PF02777; sodfe_C; 1.

DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom; PD000475; SODismutase; 2.

DR PROSITE; PS00088; SOD MN; 1.

KW Oxidoreductase; Metal-binding; Manganese; Mitochondrion;

KW Transit peptide.

FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).

```

FT CHAIN ? 200 SUPEROXIDE DISMUTASE [MN]
FT METAL 27 27 MANGANESE (BY SIMILARITY)
FT METAL 72 72 MANGANESE (BY SIMILARITY)
FT METAL 157 157 MANGANESE (BY SIMILARITY)
FT METAL 161 161 MANGANESE (BY SIMILARITY)
SQ SEQUENCE 200 AA; 22194 MW; 9758B1DD1F674F19 CRC64;

Query Match 3.3%; Score 8; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYISR 121
DB 14 ALEPYISR 21

RESULT 11
SODF_BACSU STANDARD; PRT; 281 AA.
AC Q35023;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable superoxide dismutase [rel (EC 1.15.1.1)].
GN SODF OR BSU19330.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the terC and oshAB loci cloned in a yeast artificial chromosome.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Henaut A.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Neone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror R., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 iron ion per subunit (By similarity).

```

```

CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF027868; AAB84442.1; -.
CC EMBL; Z99114; CAB13825.1; -.
CC PIR; C69709; C69709.
CC DR HSSP; P80293; 1A0M.
CC DR Subtilist; B312876; sodF.
CC DR InterPro; IPR001189; SODismutase.
CC DR Pfam; PF00081; sodfe; 1.
CC DR Pfam; PF02777; sodfe C; 1.
CC DR PRINTS; PR01703; MNSODISMUTASE.
CC DR ProDom; PD000475; SODismutase; 1.
CC DR PROSITE; PS00088; SOD_MN; 1.
CC KW Oxidoreductase; Metal-binding; Iron; Complete proteome.
FT METAL 104 104 IRON (BY SIMILARITY).
FT METAL 152 152 IRON (BY SIMILARITY).
FT METAL 236 236 IRON (BY SIMILARITY).
FT METAL 240 240 IRON (BY SIMILARITY).
SQ SEQUENCE 281 AA; 33477 MW; 7F36AC0A60E74DB0 CRC64;

Query Match 3.3%; Score 8; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYISR 121
DB 91 ALEPYISR 98

RESULT 12
CA64 HUMAN STANDARD; PRT; 1691 AA.
ID CA64 HUMAN STANDARD; PRT; 1691 AA.
AC Q14031; Q12823; Q14053; Q9N0M5; Q9NTX3; Q9UJ76; Q9UMG6; Q9Y4L4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eyes, and Kidney;
RX MEDLINE=94171779; PubMed=8125972;
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RN J. Biol. Chem. 269:7520-7526(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230419; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Reiders S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RN J. Biol. Chem. 269:13193-13199(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
RX MEDLINE=9661006; PubMed=8661006;
RA Zhang X., Zhou J., Reiders S.T., Tryggvason K.;

```

RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RL Genomics 33:473-479(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grafham D., Lawlor S., Wilson S.,
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=9361972; PubMed=8356449;
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,
RL de Paeppe A., Tryggvason K., Reenders S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169(1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in inter- and intramolecular, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D21337; BAA04809.1; -;
DR EMBL; U04845; AAA19569.2; -;
DR EMBL; U47004; AAB19038.1; -;
DR EMBL; U46959; AAB19038.1; JOINED.
DR EMBL; U46961; AAB19038.1; JOINED.
DR EMBL; U46962; AAB19038.1; JOINED.
DR EMBL; U46963; AAB19038.1; JOINED.
DR EMBL; U46964; AAB19038.1; JOINED.
DR EMBL; U46965; AAB19038.1; JOINED.
DR EMBL; U46966; AAB19038.1; JOINED.
DR EMBL; U46967; AAB19038.1; JOINED.
DR EMBL; U46968; AAB19038.1; JOINED.
DR EMBL; U46969; AAB19038.1; JOINED.
DR EMBL; U46970; AAB19038.1; JOINED.
DR EMBL; U46971; AAB19038.1; JOINED.
DR EMBL; U46972; AAB19038.1; JOINED.
DR EMBL; U46973; AAB19038.1; JOINED.
DR EMBL; U46974; AAB19038.1; JOINED.
DR EMBL; U46975; AAB19038.1; JOINED.
DR EMBL; U46976; AAB19038.1; JOINED.
DR EMBL; U46977; AAB19038.1; JOINED.
DR EMBL; U46978; AAB19038.1; JOINED.
DR EMBL; U46979; AAB19038.1; JOINED.
DR EMBL; U46980; AAB19038.1; JOINED.
DR EMBL; U46981; AAB19038.1; JOINED.
DR EMBL; U46982; AAB19038.1; JOINED.
DR EMBL; U46983; AAB19038.1; JOINED.
DR EMBL; U46984; AAB19038.1; JOINED.
DR EMBL; U46985; AAB19038.1; JOINED.
DR EMBL; U46986; AAB19038.1; JOINED.
DR EMBL; U46987; AAB19038.1; JOINED.
DR EMBL; U46988; AAB19038.1; JOINED.
DR EMBL; U46989; AAB19038.1; JOINED.
DR EMBL; U46990; AAB19038.1; JOINED.
DR EMBL; U46991; AAB19038.1; JOINED.
DR EMBL; U46992; AAB19038.1; JOINED.
DR EMBL; U46993; AAB19038.1; JOINED.
DR EMBL; U46994; AAB19038.1; JOINED.
DR EMBL; U46995; AAB19038.1; JOINED.
DR EMBL; U46996; AAB19038.1; JOINED.
DR EMBL; U46997; AAB19038.1; JOINED.
DR EMBL; U46998; AAB19038.1; JOINED.
DR EMBL; U46999; AAB19038.1; JOINED.
DR EMBL; U47000; AAB19038.1; JOINED.
DR EMBL; U47001; AAB19038.1; JOINED.
DR EMBL; U47002; AAB19038.1; JOINED.
DR EMBL; U47003; AAB19038.1; JOINED.
DR EMBL; U47004; AAB19038.1; JOINED.
DR EMBL; U47005; AAB19038.1; JOINED.
DR EMBL; U47006; AAB19038.1; JOINED.
DR EMBL; U47007; AAB19038.1; JOINED.
DR EMBL; U47008; AAB19038.1; JOINED.
DR EMBL; U47009; AAB19038.1; JOINED.
DR EMBL; U47010; AAB19038.1; JOINED.
DR EMBL; U47011; AAB19038.1; JOINED.
DR EMBL; U47012; AAB19038.1; JOINED.
DR EMBL; U47013; AAB19038.1; JOINED.
DR EMBL; U47014; AAB19038.1; JOINED.
DR EMBL; U47015; AAB19038.1; JOINED.
DR EMBL; U47016; AAB19038.1; JOINED.
DR EMBL; U47017; AAB19038.1; JOINED.
DR EMBL; U47018; AAB19038.1; JOINED.
DR EMBL; U47019; AAB19038.1; JOINED.
DR EMBL; U47020; AAB19038.1; JOINED.
DR EMBL; U47021; AAB19038.1; JOINED.
DR EMBL; U47022; AAB19038.1; JOINED.
DR EMBL; U47023; AAB19038.1; JOINED.
DR EMBL; U47024; AAB19038.1; JOINED.
DR EMBL; U47025; AAB19038.1; JOINED.
DR EMBL; U47026; AAB19038.1; JOINED.
DR EMBL; U47027; AAB19038.1; JOINED.
DR EMBL; U47028; AAB19038.1; JOINED.
DR EMBL; U47029; AAB19038.1; JOINED.
DR EMBL; U47030; AAB19038.1; JOINED.
DR EMBL; U47031; AAB19038.1; JOINED.
DR EMBL; U47032; AAB19038.1; JOINED.
DR EMBL; U47033; AAB19038.1; JOINED.
DR EMBL; U47034; AAB19038.1; JOINED.
DR EMBL; U47035; AAB19038.1; JOINED.
DR EMBL; U47036; AAB19038.1; JOINED.
DR EMBL; U47037; AAB19038.1; JOINED.
DR EMBL; U47038; AAB19038.1; JOINED.
DR EMBL; U47039; AAB19038.1; JOINED.
DR EMBL; U47040; AAB19038.1; JOINED.
DR EMBL; U47041; AAB19038.1; JOINED.
DR EMBL; U47042; AAB19038.1; JOINED.
DR EMBL; U47043; AAB19038.1; JOINED.
DR EMBL; U47044; AAB19038.1; JOINED.
DR EMBL; U47045; AAB19038.1; JOINED.
DR EMBL; U47046; AAB19038.1; JOINED.
DR EMBL; U47047; AAB19038.1; JOINED.
DR EMBL; U47048; AAB19038.1; JOINED.
DR EMBL; U47049; AAB19038.1; JOINED.
DR EMBL; U47050; AAB19038.1; JOINED.
DR EMBL; U47051; AAB19038.1; JOINED.
DR EMBL; U47052; AAB19038.1; JOINED.
DR EMBL; U47053; AAB19038.1; JOINED.
DR EMBL; U47054; AAB19038.1; JOINED.
DR EMBL; U47055; AAB19038.1; JOINED.
DR EMBL; U47056; AAB19038.1; JOINED.
DR EMBL; U47057; AAB19038.1; JOINED.
DR EMBL; U47058; AAB19038.1; JOINED.
DR EMBL; U47059; AAB19038.1; JOINED.
DR EMBL; U47060; AAB19038.1; JOINED.
DR EMBL; U47061; AAB19038.1; JOINED.
DR EMBL; U47062; AAB19038.1; JOINED.
DR EMBL; U47063; AAB19038.1; JOINED.
DR EMBL; U47064; AAB19038.1; JOINED.
DR EMBL; U47065; AAB19038.1; JOINED.
DR EMBL; U47066; AAB19038.1; JOINED.
DR EMBL; U47067; AAB19038.1; JOINED.
DR EMBL; U47068; AAB19038.1; JOINED.
DR EMBL; U47069; AAB19038.1; JOINED.
DR EMBL; U47070; AAB19038.1; JOINED.
DR EMBL; U47071; AAB19038.1; JOINED.
DR EMBL; U47072; AAB19038.1; JOINED.
DR EMBL; U47073; AAB19038.1; JOINED.
DR EMBL; U47074; AAB19038.1; JOINED.
DR EMBL; U47075; AAB19038.1; JOINED.
DR EMBL; U47076; AAB19038.1; JOINED.
DR EMBL; U47077; AAB19038.1; JOINED.
DR EMBL; U47078; AAB19038.1; JOINED.
DR EMBL; U47079; AAB19038.1; JOINED.
DR EMBL; U47080; AAB19038.1; JOINED.
DR EMBL; U47081; AAB19038.1; JOINED.
DR EMBL; U47082; AAB19038.1; JOINED.
DR EMBL; U47083; AAB19038.1; JOINED.
DR EMBL; U47084; AAB19038.1; JOINED.
DR EMBL; U47085; AAB19038.1; JOINED.
DR EMBL; U47086; AAB19038.1; JOINED.
DR EMBL; U47087; AAB19038.1; JOINED.
DR EMBL; U47088; AAB19038.1; JOINED.
DR EMBL; U47089; AAB19038.1; JOINED.
DR EMBL; U47090; AAB19038.1; JOINED.
DR EMBL; U47091; AAB19038.1; JOINED.
DR EMBL; U47092; AAB19038.1; JOINED.
DR EMBL; U47093; AAB19038.1; JOINED.
DR EMBL; U47094; AAB19038.1; JOINED.
DR EMBL; U47095; AAB19038.1; JOINED.
DR EMBL; U47096; AAB19038.1; JOINED.
DR EMBL; U47097; AAB19038.1; JOINED.
DR EMBL; U47098; AAB19038.1; JOINED.
DR EMBL; U47099; AAB19038.1; JOINED.
DR EMBL; U47100; AAB19038.1; JOINED.
DR EMBL; U47101; AAB19038.1; JOINED.
DR EMBL; U47102; AAB19038.1; JOINED.
DR EMBL; U47103; AAB19038.1; JOINED.
DR EMBL; U47104; AAB19038.1; JOINED.
DR EMBL; U47105; AAB19038.1; JOINED.
DR EMBL; U47106; AAB19038.1; JOINED.
DR EMBL; U47107; AAB19038.1; JOINED.
DR EMBL; U47108; AAB19038.1; JOINED.
DR EMBL; U47109; AAB19038.1; JOINED.
DR EMBL; U47110; AAB19038.1; JOINED.
DR EMBL; U47111; AAB19038.1; JOINED.
DR EMBL; U47112; AAB19038.1; JOINED.
DR EMBL; U47113; AAB19038.1; JOINED.
DR EMBL; U47114; AAB19038.1; JOINED.
DR EMBL; U47115; AAB19038.1; JOINED.
DR EMBL; U47116; AAB19038.1; JOINED.
DR EMBL; U47117; AAB19038.1; JOINED.
DR EMBL; U47118; AAB19038.1; JOINED.
DR EMBL; U47119; AAB19038.1; JOINED.
DR EMBL; U47120; AAB19038.1; JOINED.
DR EMBL; U47121; AAB19038.1; JOINED.
DR EMBL; U47122; AAB19038.1; JOINED.
DR EMBL; U47123; AAB19038.1; JOINED.
DR EMBL; U47124; AAB19038.1; JOINED.
DR EMBL; U47125; AAB19038.1; JOINED.
DR EMBL; U47126; AAB19038.1; JOINED.
DR EMBL; U47127; AAB19038.1; JOINED.
DR EMBL; U47128; AAB19038.1; JOINED.
DR EMBL; U47129; AAB19038.1; JOINED.
DR EMBL; U47130; AAB19038.1; JOINED.
DR EMBL; U47131; AAB19038.1; JOINED.
DR EMBL; U47132; AAB19038.1; JOINED.
DR EMBL; U47133; AAB19038.1; JOINED.
DR EMBL; U47134; AAB19038.1; JOINED.
DR EMBL; U47135; AAB19038.1; JOINED.
DR EMBL; U47136; AAB19038.1; JOINED.
DR EMBL; U47137; AAB19038.1; JOINED.
DR EMBL; U47138; AAB19038.1; JOINED.
DR EMBL; U47139; AAB19038.1; JOINED.
DR EMBL; U47140; AAB19038.1; JOINED.
DR EMBL; U47141; AAB19038.1; JOINED.
DR EMBL; U47142; AAB19038.1; JOINED.
DR EMBL; U47143; AAB19038.1; JOINED.
DR EMBL; U47144; AAB19038.1; JOINED.
DR EMBL; U47145; AAB19038.1; JOINED.
DR EMBL; U47146; AAB19038.1; JOINED.
DR EMBL; U47147; AAB19038.1; JOINED.
DR EMBL; U47148; AAB19038.1; JOINED.
DR EMBL; U47149; AAB19038.1; JOINED.
DR EMBL; U47150; AAB19038.1; JOINED.
DR EMBL; U47151; AAB19038.1; JOINED.
DR EMBL; U47152; AAB19038.1; JOINED.
DR EMBL; U47153; AAB19038.1; JOINED.
DR EMBL; U47154; AAB19038.1; JOINED.
DR EMBL; U47155; AAB19038.1; JOINED.
DR EMBL; U47156; AAB19038.1; JOINED.
DR EMBL; U47157; AAB19038.1; JOINED.
DR EMBL; U47158; AAB19038.1; JOINED.
DR EMBL; U47159; AAB19038.1; JOINED.
DR EMBL; U47160; AAB19038.1; JOINED.
DR EMBL; U47161; AAB19038.1; JOINED.
DR EMBL; U47162; AAB19038.1; JOINED.
DR EMBL; U47163; AAB19038.1; JOINED.
DR EMBL; U47164; AAB19038.1; JOINED.
DR EMBL; U47165; AAB19038.1; JOINED.
DR EMBL; U47166; AAB19038.1; JOINED.
DR EMBL; U47167; AAB19038.1; JOINED.
DR EMBL; U47168; AAB19038.1; JOINED.
DR EMBL; U47169; AAB19038.1; JOINED.
DR EMBL; U47170; AAB19038.1; JOINED.
DR EMBL; U47171; AAB19038.1; JOINED.
DR EMBL; U47172; AAB19038.1; JOINED.
DR EMBL; U47173; AAB19038.1; JOINED.
DR EMBL; U47174; AAB19038.1; JOINED.
DR EMBL; U47175; AAB19038.1; JOINED.
DR EMBL; U47176; AAB19038.1; JOINED.
DR EMBL; U47177; AAB19038.1; JOINED.
DR EMBL; U47178; AAB19038.1; JOINED.
DR EMBL; U47179; AAB19038.1; JOINED.
DR EMBL; U47180; AAB19038.1; JOINED.
DR EMBL; U47181; AAB19038.1; JOINED.
DR EMBL; U47182; AAB19038.1; JOINED.
DR EMBL; U47183; AAB19038.1; JOINED.
DR EMBL; U47184; AAB19038.1; JOINED.
DR EMBL; U47185; AAB19038.1; JOINED.
DR EMBL; U47186; AAB19038.1; JOINED.
DR EMBL; U47187; AAB19038.1; JOINED.
DR EMBL; U47188; AAB19038.1; JOINED.
DR EMBL; U47189; AAB19038.1; JOINED.
DR EMBL; U47190; AAB19038.1; JOINED.
DR EMBL; U47191; AAB19038.1; JOINED.
DR EMBL; U47192; AAB19038.1; JOINED.
DR EMBL; U47193; AAB19038.1; JOINED.
DR EMBL; U47194; AAB19038.1; JOINED.
DR EMBL; U47195; AAB19038.1; JOINED.
DR EMBL; U47196; AAB19038.1; JOINED.
DR EMBL; U47197; AAB19038.1; JOINED.
DR EMBL; U47198; AAB19038.1; JOINED.
DR EMBL; U47199; AAB19038.1; JOINED.
DR EMBL; U47200; AAB19038.1; JOINED.
DR EMBL; U47201; AAB19038.1; JOINED.
DR EMBL; U47202; AAB19038.1; JOINED.
DR EMBL; U47203; AAB19038.1; JOINED.
DR EMBL; U47204; AAB19038.1; JOINED.
DR EMBL; U47205; AAB19038.1; JOINED.
DR EMBL; U47206; AAB19038.1; JOINED.
DR EMBL; U47207; AAB19038.1; JOINED.
DR EMBL; U47208; AAB19038.1; JOINED.
DR EMBL; U47209; AAB19038.1; JOINED.
DR EMBL; U47210; AAB19038.1; JOINED.
DR EMBL; U47211; AAB19038.1; JOINED.
DR EMBL; U47212; AAB19038.1; JOINED.
DR EMBL; U47213; AAB19038.1; JOINED.
DR EMBL; U47214; AAB19038.1; JOINED.
DR EMBL; U47215; AAB19038.1; JOINED.
DR EMBL; U47216; AAB19038.1; JOINED.
DR EMBL; U47217; AAB19038.1; JOINED.
DR EMBL; U47218; AAB19038.1; JOINED.
DR EMBL; U47219; AAB19038.1; JOINED.
DR EMBL; U47220; AAB19038.1; JOINED.
DR EMBL; U47221; AAB19038.1; JOINED.
DR EMBL; U47222; AAB19038.1; JOINED.
DR EMBL; U47223; AAB19038.1; JOINED.
DR EMBL; U47224; AAB19038.1; JOINED.
DR EMBL; U47225; AAB19038.1; JOINED.
DR EMBL; U47226; AAB19038.1; JOINED.
DR EMBL; U47227; AAB19038.1; JOINED.
DR EMBL; U47228; AAB19038.1; JOINED.
DR EMBL; U47229; AAB19038.1; JOINED.
DR EMBL; U47230; AAB19038.1; JOINED.
DR EMBL; U47231; AAB19038.1; JOINED.
DR EMBL; U47232; AAB19038.1; JOINED.
DR EMBL; U47233; AAB19038.1; JOINED.
DR EMBL; U47234; AAB19038.1; JOINED.
DR EMBL; U47235; AAB19038.1; JOINED.
DR EMBL; U47236; AAB19038.1; JOINED.
DR EMBL; U47237; AAB19038.1; JOINED.
DR EMBL; U47238; AAB19038.1; JOINED.
DR EMBL; U47239; AAB19038.1; JOINED.
DR EMBL; U47240; AAB19038.1; JOINED.
DR EMBL; U47241; AAB19038.1; JOINED.
DR EMBL; U47242; AAB19038.1; JOINED.
DR EMBL; U47243; AAB19038.1; JOINED.
DR EMBL; U47244; AAB19038.1; JOINED.
DR EMBL; U47245; AAB19038.1; JOINED.
DR EMBL; U47246; AAB19038.1; JOINED.
DR EMBL; U47247; AAB19038.1; JOINED.
DR EMBL; U47248; AAB19038.1; JOINED.
DR EMBL; U47249; AAB19038.1; JOINED.
DR EMBL; U47250; AAB19038.1; JOINED.
DR EMBL; U47251; AAB19038.1; JOINED.
DR EMBL; U47252; AAB19038.1; JOINED.
DR EMBL; U47253; AAB19038.1; JOINED.
DR EMBL; U47254; AAB19038.1; JOINED.
DR EMBL; U47255; AAB19038.1; JOINED.
DR EMBL; U47256; AAB19038.1; JOINED.
DR EMBL; U47257; AAB19038.1; JOINED.
DR EMBL; U47258; AAB19038.1; JOINED.
DR EMBL; U47259; AAB19038.1; JOINED.
DR EMBL; U47260; AAB19038.1; JOINED.
DR EMBL; U47261; AAB19038.1; JOINED.
DR EMBL; U47262; AAB19038.1; JOINED.
DR EMBL; U47263; AAB19038.1; JOINED.
DR EMBL; U47264; AAB19038.1; JOINED.
DR EMBL; U47265; AAB19038.1; JOINED.
DR EMBL; U47266; AAB19038.1; JOINED.
DR EMBL; U47267; AAB19038.1; JOINED.
DR EMBL; U47268; AAB19038.1; JOINED.
DR EMBL; U47269; AAB19038.1; JOINED.
DR EMBL; U47270; AAB19038.1; JOINED.
DR EMBL; U47271; AAB19038.1; JOINED.
DR EMBL; U47272; AAB19038.1; JOINED.
DR EMBL; U47273; AAB19038.1; JOINED.
DR EMBL; U47274; AAB19038.1; JOINED.
DR EMBL; U47275; AAB19038.1; JOINED.
DR EMBL; U47276; AAB19038.1; JOINED.
DR EMBL; U47277; AAB19038.1; JOINED.
DR EMBL; U47278; AAB19038.1; JOINED.
DR EMBL; U47279; AAB19038.1; JOINED.
DR EMBL; U47280; AAB19038.1; JOINED.
DR EMBL; U47281; AAB19038.1; JOINED.
DR EMBL; U47282; AAB19038.1; JOINED.
DR EMBL; U47283; AAB19038.1; JOINED.
DR EMBL; U47284; AAB19038.1; JOINED.
DR EMBL; U47285; AAB19038.1; JOINED.
DR EMBL; U47286; AAB19038.1; JOINED.
DR EMBL; U47287; AAB19038.1; JOINED.
DR EMBL; U47288; AAB19038.1; JOINED.
DR EMBL; U47289; AAB19038.1; JOINED.
DR EMBL; U47290; AAB19038.1; JOINED.
DR EMBL; U47291; AAB19038.1; JOINED.
DR EMBL; U47292; AAB19038.1; JOINED.
DR EMBL; U47293; AAB19038.1; JOINED.
DR EMBL; U47294; AAB19038.1; JOINED.
DR EMBL; U47295; AAB19038.1; JOINED.
DR EMBL; U47296; AAB19038.1; JOINED.
DR EMBL; U47297; AAB19038.1; JOINED.
DR EMBL; U47298; AAB19038.1; JOINED.
DR EMBL; U47299; AAB19038.1; JOINED.
DR EMBL; U47300; AAB19038.1; JOINED.
DR EMBL; U47301; AAB19038.1; JOINED.
DR EMBL; U47302; AAB19038.1; JOINED.
DR EMBL; U47303; AAB19038.1; JOINED.
DR EMBL; U47304; AAB19038.1; JOINED.
DR EMBL; U47305; AAB19038.1; JOINED.
DR EMBL; U47306; AAB19038.1; JOINED.
DR EMBL; U47307; AAB19038.1; JOINED.
DR EMBL; U47308; AAB19038.1; JOINED.
DR EMBL; U47309; AAB19038.1; JOINED.
DR EMBL; U47310; AAB19038.1; JOINED.
DR EMBL; U47311; AAB19038.1; JOINED.
DR EMBL; U47312; AAB19038.1; JOINED.
DR EMBL; U47313; AAB19038.1; JOINED.
DR EMBL; U47314; AAB19038.1; JOINED.
DR EMBL; U47315; AAB19038.1; JOINED.
DR EMBL; U47316; AAB19038.1; JOINED.
DR EMBL; U47317; AAB19038.1; JOINED.
DR EMBL; U47318; AAB19038.1; JOINED.
DR EMBL; U47319; AAB19038.1; JOINED.
DR EMBL; U47320; AAB19038.1; JOINED.
DR EMBL; U47321; AAB19038.1; JOINED.
DR EMBL; U47322; AAB19038.1; JOINED.
DR EMBL; U47323; AAB19038.1; JOINED.
DR EMBL; U47324; AAB19038.1; JOINED.
DR EMBL; U47325; AAB19038.1; JOINED.
DR EMBL; U47326; AAB19038.1; JOINED.
DR EMBL; U47327; AAB19038.1; JOINED.
DR EMBL; U47328; AAB19038.1; JOINED.
DR EMBL; U47329; AAB19038.1; JOINED.
DR EMBL; U47330; AAB19038.1; JOINED.
DR EMBL; U47331; AAB19038.1; JOINED.
DR EMBL; U47332; AAB19038.1; JOINED.
DR EMBL; U47333; AAB19038.1; JOINED.
DR EMBL; U47334; AAB19038.1; JOINED.
DR EMBL; U47335; AAB19038.1; JOINED.
DR EMBL; U47336; AAB19038.1; JOINED.
DR EMBL; U47337; AAB19038.1; JOINED.
DR EMBL; U47338; AAB19038.1; JOINED.
DR EMBL; U47339; AAB19038.1; JOINED.
DR EMBL; U47340; AAB19038.1; JOINED.
DR EMBL; U47341; AAB19038.1; JOINED.
DR EMBL; U47342; AAB19038.1; JOINED.
DR EMBL; U47343; AAB19038.1; JOINED.
DR EMBL; U47344; AAB19038.1; JOINED.
DR EMBL; U47345; AAB19038.1; JOINED.
DR EMBL; U47346; AAB19038.1; JOINED.
DR EMBL; U47347; AAB19038.1; JOINED.
DR EMBL; U47348; AAB19038.1; JOINED.
DR EMBL; U47349; AAB19038.1; JOINED.
DR EMBL; U47350; AAB19038.1; JOINED.
DR EMBL; U47351; AAB19038.1; JOINED.
DR EMBL; U47352; AAB19038.1; JOINED.
DR EMBL; U47353; AAB19038.1; JOINED.
DR EMBL; U47354; AAB19038.1; JOINED.
DR EMBL; U47355; AAB19038.1; JOINED.
DR EMBL; U47356; AAB19038.1; JOINED.
DR EMBL; U47357; AAB19038.1; JOINED.
DR EMBL; U47358; AAB19038.1; JOINED.
DR EMBL; U47359; AAB19038.1; JOINED.
DR EMBL; U47360; AAB19038.1; JOINED.
DR EMBL; U47361; AAB19038.1; JOINED.
DR EMBL; U47362; AAB19038.1; JOINED.
DR EMBL; U47363; AAB19038.1; JOINED.
DR EMBL; U47364; AAB19038.1; JOINED.
DR EMBL; U47365; AAB19038.1; JOINED.
DR EMBL; U47366; AAB19038.1; JOINED.
DR EMBL; U47367; AAB19038.1; JOINED.
DR EMBL; U47368; AAB19038.1; JOINED.
DR EMBL; U47369; AAB19038.1; JOINED.
DR EMBL; U47370; AAB19038.1; JOINED.
DR EMBL; U47371; AAB19038.1; JOINED.
DR EMBL; U47372; AAB19038.1; JOINED.
DR EMBL; U47373; AAB19038.1; JOINED.
DR EMBL; U47374; AAB19038.1; JOINED.
DR EMBL; U47375; AAB19038.1; JOINED.
DR EMBL; U47376; AAB19038.1; JOINED.
DR EMBL; U47377; A

DR EMBL; L22763; AAL16338.1; --
DR PIR; A54122; CGHUGB
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; --
DR GO; GO:0005587; C:collagen type IV; NAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR SMART; SM003923; ProcollagnC4; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR ProDom; PD000007; Clg helix; 9.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR SMART; SM003923; ProcollagnC4; 1.
DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
DR Signal; 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6 (IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6 (IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.
Query Match 3.3%; Score 8; DB 1; Length 1691;
Best/Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 234 SRCQVCMK 241
Db 1682 SRCQVCMK 1689
RESULT 13
CA14 DROME STANDARD; PRT; 1775 AA.
AC P08120;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN CG25C OR DCG1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89054012; PubMed=3142875;
RA Blumberg B., Mackrell A.J., Fessler J.H.;
RT "Drosophila basement membrane procollagen alpha 1(IV). II. Complete
RT cDNA sequence, genomic structure, and general implications for
RT supramolecular assemblies."
RL J. Biol. Chem. 263:18328-18337(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87194801; PubMed=3106346;
RA Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,
RA Natzie J.E., Fessler J.H.;
RT "Basement membrane procollagen IV and its specialized carboxyl domain
RT are conserved in Drosophila, mouse, and human."
RL J. Biol. Chem. 262:5947-5950(1987).
RN [5]
RP SEQUENCE OF 1355-1775 FROM N.A.
RX MEDLINE=87246644; PubMed=3109906;
RA Cecchini J.P., Knibiehler B., Mirre C., le Parco Y.;
"Evidence for a type-IV-related collagen in Drosophila melanogaster.
Evolutionary constancy of the carboxyl-terminal noncollagenous
domain.";
Eur. J. Biochem. 165:587-593(1987).
[6]
RP SEQUENCE OF 762-1230 FROM N.A.
RX MEDLINE=82197577; PubMed=6210912;
RA Monson J.M., Natzie J., Friedman J., McCarthy B.J.;
RA "Expression and novel structure of a collagen gene in Drosophila.";
Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).
CC -|- FUNCTION: Collagen type IV is specific for basement membranes.
CC -|- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -|- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -|- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -|- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; M23704; AAA28404.1; --
EMBL; M96575; AAB55184.1; --
EMBL; J02727; AAA28423.1; --
EMBL; M28334; AAA28422.1; --
EMBL; V00200; CAA23486.2; --
PIR; A31893; A31893.
FlyBase; FBgn000299; Cg25C.
GO; GO:0005587; C:collagen type IV; NAS.
InterPro; IPR008161; Clg helix.
InterPro; IPR008160; Collagen.
InterPro; IPR001442; Procollagn4_C.
Pfam; PF01413; C4; 2.
Pfam; PF01391; Collagen; 25.
ProDom; PD000007; Clg helix; 9.
ProDom; PD003923; ProcollagnC4; 1.
SMART; SM00111; C4; 2.
SMART; SM003923; ProcollagnC4; 1.
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
SIGNAL 1 23
FT PROPEP 24 ? AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN ? 1775 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN ? 1544 TRIPLE-HELICAL REGION.
FT DOMAIN 1545 1775 NONHELICAL REGION (NC1).
FT DISULFID 1569 1655 OR 1652 (BY SIMILARITY).
FT DISULFID 1599 1652 OR 1655 (BY SIMILARITY).
FT DISULFID 1611 1617 BY SIMILARITY.
FT DISULFID 1674 1770 OR 1767 (BY SIMILARITY).
FT DISULFID 1708 1767 OR 1770 (BY SIMILARITY).
FT DISULFID 1720 1727 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 948 948 L -> S (IN REF. 6).
FT CONFLICT 997 997 S -> T (IN REF. 6).
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).
FT CONFLICT 1373 1373 T -> I (IN REF. 5).
FT CONFLICT 1496 1496 L -> R (IN REF. 5).
FT CONFLICT 1507 1511 ETCNV -> RAGOR (IN REF. 5).
FT CONFLICT 1529 1529 E -> K (IN REF. 5).


```
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AE23149525CD CRC64;

Query Match 3.3%; Score 8; DB 1; Length 1775;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
DB 1765 SRCQVCWK 1772
|||||

RESULT 14
VB03 VACCC STANDARD; PRT; 124 AA.
AC P21050;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein B3.
GN B3R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP004599; BAC13963.1; -
DR HAMAP; MF_00651; -; 1.
DR InterPro; IPR005227; Cons_hypoth250.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF03652; UPF0081; 1.
DR SMART; SMC0732; YqgFc; 1.
DR TIGRPFAMs; TIGR00250; TIGR00250; 1.
DR Hydrolase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ SEQUENCE 137 AA; 15239 MW; 500BDADCFCBDB7CEC CRC64;

Query Match 2.9%; Score 7; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIS 234
DB 44 ELEKIIS 50
|||||

RESULT 16
RNB HSV2H STANDARD; PRT; 151 AA.
AC P89479;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Potential RNA-binding protein.
GN US11.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC
CC -!- FUNCTION: BINDS DNA AND RNA (BY SIMILARITY).
CC
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z86099; CAB06719.1; -
DR DNA-binding; RNA-binding; Repeat; Nuclear protein.
FT DOMAIN 90 146 11 X 6 AA TANDEM REPEATS.
FT REPEAT 90 95 1.
FT REPEAT 96 101 2.
```

FT REPEAT 102 104 3.
FT REPEAT 105 110 4.
FT REPEAT 111 116 5.
FT REPEAT 117 122 6.
FT REPEAT 123 128 7.
FT REPEAT 129 130 8.
FT REPEAT 131 134 9.
FT REPEAT 135 140 10.
FT REPEAT 141 146 11.
SQ SEQUENCE 151 AA; 16297 MW; FAB751F23C3DB6AE CRC64;

Query Match 2.9%; Score 7; DB 1; Length 151;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GKRGSQ 10
|||||
DB 61 GKRGSQ 67

RESULT 17

ID NUPM_NEUCR STANDARD; PRT; 183 AA.
AC P21976;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE NADH-ubiquinone oxidoreductase 20.8 kDa subunit (EC 1.6.5.3)
DE (EC 1.6.99.3).
OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-45.
RA MEDLINE=90330847; PubMed=2142943;

RX Videira A., Tropisch M., Wachter E., Schneider H., Werner S.;
RT "Molecular cloning of subunits of complex I from Neurospora crassa.
RT Primary structure and in vitro expression of a 22-kDa polypeptide.";

RL J. Biol. Chem. 265:13060-13065 (1990).

CC -!- FUNCTION: Transfer of electrons from NADH to the respiratory chain. The immediate electron acceptor for the enzyme is believed to be ubiquinone.
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC -!- COFACTOR: Binds 1 iron-sulfur cluster (potential).

CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.

CC -!- THIS IS A COMPONENT OF THE HYDROPHOBIC FRACTION.

CC -!- SUBCELLULAR LOCATION: Matrix and cytoplasmic side of the mitochondrial inner membrane.

CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 19 kDa SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55323; AAA33571.1; -
CC PIR; T47251; T47251.
CC InterPro; IPR008697; NDUFA8.
CC Pfam; PF05950; NDUFA8; 1.
CC KX Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 183 AA; 20911 MW; A2574693F41093D4 CRC64;

QY 177 SCLEEF 183
|||||

Query Match 2.9%; Score 7; DB 1; Length 183;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 90 SCLEEF 96

RESULT 18

ID SODM_PROFR STANDARD; PRT; 201 AA.
AC P80233;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).
GN SODA.

OS Propionibacterium freudenreichii shtermanii.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.

OX NCBI_TaxID=1752;

RN [1]

RP SEQUENCE.

RC STRAIN=P23;

RX MEDLINE=94139724; PubMed=8307013;

RA Meier B., Sehn A.P., Schinina M.E., Barra D.;

RT "In vivo incorporation of copper into the iron-exchangeable and

RT manganese-exchangeable superoxide dismutase from Propionibacterium

RT shtermanii. Amino acid sequence and identity of the protein

RT moieties.";

RL Eur. J. Biochem. 219:463-468 (1994).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RA Schmidt M., Meier B., Parak F.;

RT "X-ray structure of the cambialistic superoxide dismutase from

RT Propionibacterium shtermanii active with Fe or Mn.";

RL J. Biol. Inorg. Chem. 1:532-541 (1996).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).

RA Schmidt M., Scherck C., Iakovleva O., Nolting H.F., Meier B., Parak F.;

RT Submitted (SEP-1997) to the PDB data bank.

RN [4]

RX MEDLINE=99248073; PubMed=10231372;

RA Schmidt M.;

RT "Manipulating the coordination number of the ferric iron within the

RT cambialistic superoxide dismutase of Propionibacterium shtermanii by

RT changing the pH-value. A crystallographic analysis.";

RL Eur. J. Biochem. 262:117-127 (1999).

CC -!- FUNCTION: Destroys radicals which are normally produced within the

CC cells and which are toxic to biological systems.

CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -!- COFACTOR: Binds 1 manganese or iron ion per subunit.

CC -!- SUBUNIT: Homotetramer.

CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase

CC family.

DR PIR; JC4396; JC4396.

DR PDB; 1AR4; 12-NOV-97.

DR PDB; 1ARS; 12-NOV-97.

DR PDB; 1AVN; 18-MAR-98.

DR PDB; 1BS3; 15-JUN-99.

DR PDB; 1BSM; 15-JUN-99.

DR PDB; 1B78; 15-JUN-99.

DR InterPro; IPR001189; SODismutase.

DR Pfam; PF00081; so_dfe; 1.

DR Pfam; PF02777; so_dfe; 1.

DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom; PD000475; SODismutase; 1.

DR PROSITE; PS00088; SOD MN; 1.

KW Oxidoreductase; Metal-binding; Manganese; Iron; 3D-structure.

FT METAL 27 27

FT METAL 75 75

FT METAL 161 161

FT METAL 165 165

FT TURN 12 18

FT HELIX 21 29

FT TURN 30 30

FT HELIX 31 52

```
FT TURN 53 53
FT TURN 56 57
FT HELIX 58 80
FT TURN 81 81
FT STRAND 82 82
FT TURN 85 86
FT HELIX 94 104
FT HELIX 107 119
FT TURN 120 120
FT STRAND 125 132
FT TURN 133 136
FT STRAND 137 144
FT TURN 145 147
FT STRAND 148 148
FT TURN 152 153
FT STRAND 155 161
FT HELIX 164 166
FT TURN 167 167
FT HELIX 168 171
FT TURN 172 173
FT HELIX 175 182
FT STRAND 183 185
FT HELIX 186 186
FT TURN 188 198
FT TURN 199 199
SQ SEQUENCE 201 AA; 22633 MW; 5BFEF424C7B32E00 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 201;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 14 ALEPYIS 20

RESULT 19
SODF_METTM STANDARD; PRT; 202 AA.
AC Q60036;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD.
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=79929;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95301176; PubMed=7781971;
RA Meile L., Fischer K., Leisinger T.;
RT "Characterization of the superoxide dismutase gene and its upstream
region from Methanobacterium thermoautotrophicum Marburg.";
RL FEMS Microbiol. Lett. 128:247-253 (1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit (By similarity).
CC -!- SUBUNIT: Homotetramer.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X74264; CAA52323.1; -.
```

```
DR PIR; S51097; S51097.
DR HSSP; Q08713; 1B06.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe.C; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR ProDom; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase; Metal-binding; Iron.
FT METAL 30 30 IRON (BY SIMILARITY).
FT METAL 78 78 IRON (BY SIMILARITY).
FT METAL 164 164 IRON (BY SIMILARITY).
FT METAL 168 168 IRON (BY SIMILARITY).
SQ SEQUENCE 202 AA; 23829 MW; SC4FBE27EE63223 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 17 ALEPYIS 23

RESULT 20
SODF_SULAC STANDARD; PRT; 210 AA.
AC Q08713;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93326644; PubMed=8334170;
RA Klenk H.-P., Schleper C., Schwass V., Brudler R.;
RT "Nucleotide sequence, transcription and phylogeny of the gene
encoding the superoxide dismutase of Sulfolobus acidocaldarius.";
RL Biochim. Biophys. Acta 1174:95-98 (1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=99096955; PubMed=9878438;
RA Knapp S., Kardinahl S., Hellgren N., Tibbelin G., Schaefer G.,
RA Ladenstein R.;
RT "Refined crystal structure of a superoxide dismutase from the
hyperthermophilic archaeon Sulfolobus acidocaldarius at 2.2-A
resolution.";
RL J. Mol. Biol. 285:689-702 (1999).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit.
CC -!- SUBUNIT: HOMOTETRAMER AT HIGH TEMPERATURE; HOMODIMER AT ROOM
TEMPERATURE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X63386; CAA44993.1; -.
```

```

DR PIR; S34616; S34616.
DR PDB; 1B06; 18-NOV-99.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; FALSE NEG.
KW Oxidoreductase; Metal-binding; Iron; 3D-structure.
FT INIT_MET 0
FT METAL 33
FT METAL 33 IRON.
FT METAL 84
FT METAL 84 IRON.
FT METAL 170
FT METAL 170 IRON.
FT METAL 174
FT METAL 174 IRON.
FT TURN 18
FT TURN 24
FT HELIX 27
FT TURN 35
FT TURN 35
FT TURN 36
FT HELIX 37
FT TURN 58
FT TURN 62
FT TURN 63
FT HELIX 67
FT TURN 89
FT TURN 90
FT TURN 91
FT STRAND 91
FT TURN 94
FT HELIX 103
FT HELIX 113
FT STRAND 116
FT STRAND 129
FT TURN 134
FT TURN 142
FT STRAND 147
FT STRAND 154
FT TURN 156
FT STRAND 157
FT TURN 161
FT STRAND 165
FT HELIX 173
FT TURN 176
FT TURN 177
FT HELIX 181
FT TURN 182
FT HELIX 184
FT HELIX 191
FT STRAND 192
FT STRAND 195
FT HELIX 197
FT TURN 205
FT TURN 206
SQ SEQUENCE 210 AA; 24135 MW; 086CCAB277D99FBB CRC64;
Query Match 2.9%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 114 ALEPYIS 120
Db 20 ALEPYIS 26
|||||
RESULT 21.
SODF_SULSO STANDARD; PRT; 210 AA.
AC P00837;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD OR SSO0316.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE.
RC STRAIN=ATCC 49255 / DSM 5833 / MT-4;
RX MEDLINE=98088931; PubMed=9428655;
RA Dello Russo A., Rullo R., Nitti G., Masullo M., Bocchini V.;
RT "Iron superoxide dismutase from the archaeon Sulfolobus solfataricus:
average hydrophobicity and amino acid weight are involved in the
adaptation of proteins to extreme environments.";
Biochim. Biophys. Acta 1343:23-30(1997).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99098843; PubMed=9880816;
RA Yamano S., Maruyama T.;
RT "An azide-insensitive superoxide dismutase from a hyperthermophilic
archaeon, Sulfolobus solfataricus.";
J. Biochem. 125:186-193(1999).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49255 / DSM 5833 / MT-4;
RX MEDLINE=21145482; PubMed=11248699;
RA De Vendittis E., Ursby T., Rullo R., Gogliettino M.A., Masullo M.,
Bocchini V.;
RT "Phenylmethanesulfonyl fluoride inactivates an archaeal superoxide
dismutase by chemical modification of a specific tyrosine residue:
cloning, sequencing and expression of the gene coding for Sulfolobus
solfataricus superoxide dismutase.";
Eur. J. Biochem. 268:1794-1801(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
[5]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=99134399; PubMed=9931259;
RA Urby T., Adinolfi B.S., Al-Karadaghi S., de Vendittis E.,
Bocchini V.;
RT "Iron superoxide dismutase from the archaeon Sulfolobus solfataricus:
analysis of structure and thermostability.";
J. Mol. Biol. 286:189-205(1999).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB012620; BAA75509.1; -
EMBL; Y15326; CAA75583.1; -
EMBL; AE006666; AAK40652.1; -
DR PIR; E90174; E90174.
DR PDB; 1SSS; 09-APR-99.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; FALSE NEG.
KW Oxidoreductase; Metal-binding; Iron; 3D-structure; Complete proteome.
FT INIT_MET 0
FT METAL 37
FT METAL 37 IRON.
FT METAL 84
FT METAL 84 IRON.

```

```

FT METAL 170 170 IRON.
FT METAL 174 174 IRON.
FT TURN 18 24
FT HELIX 27 35
FT TURN 36 36
FT HELIX 37 57
FT TURN 58 58
FT TURN 62 63
FT HELIX 67 89
FT TURN 90 90
FT STRAND 91 91
FT STRAND 94 94
FT TURN 95 96
FT STRAND 97 97
FT HELIX 103 113
FT HELIX 116 129
FT STRAND 134 140
FT TURN 142 144
FT STRAND 147 153
FT TURN 154 156
FT STRAND 157 157
FT STRAND 165 170
FT HELIX 173 175
FT TURN 176 176
FT HELIX 177 180
FT TURN 181 182
FT HELIX 184 191
FT HELIX 192 194
FT STRAND 195 195
FT HELIX 197 206
FT TURN 207 207
SQ SEQUENCE 210 AA; 24112 MW; 7918CF1292BF98B6 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 114 ALEPYIS 120
    |||||
Db 20 ALEPYIS 26

RESULT 22
SODF ACIAM STANDARD; PRT; 211 AA.
AC Q9P9L3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD.
OS Acidianus ambivalens (Desulfurolobus ambivalens).
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Acidianus.
OX NCBI_TaxID=2283;
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE=21026956; PubMed=11154067;
RA Kardinahl S., Anemuller S., Schaefer G.;
RT "The hyper-thermostable Fe-superoxide dismutase from the Archaeon
RT Acidianus ambivalens: characterization, recombinant expression,
RT crystallization and effects of metal exchange.";
RL Biol. Chem. 381:1089-1101(2000).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```
DR HSP: Q08713; 1B06
DR InterPro: IPR001189; SODismutase.
DR Pfam: PF00081; sodfe; 1.
DR PRINTS: PF02777; sodfe; 1.
DR PRODOM: PD000475; SODismutase; 1.
DR PROSITE: PS00088; SOD MN; 1.
KW Oxidoreductase; Metal-binding; Iron; Complete proteome.
FT METAL 31 31 IRON (BY SIMILARITY).
FT METAL 79 79 IRON (BY SIMILARITY).
FT METAL 165 165 IRON (BY SIMILARITY).
FT METAL 169 169 IRON (BY SIMILARITY).
SQ SEQUENCE 211 AA; 24204 MW; F5943D397DD31561 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 18 ALEPYIS 24

RESULT 24
SODM_YEAST
ID SODM_YEAST STANDARD; PRT; 233 AA.
AC P00447;
DT 21-JUL-1986 (Rel. 01, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
GN SOD2 OR YHR008C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
[2]
SEQUENCE FROM N.A.
RA Martes C.A.M., van Loon A.P.G.M., Oudshoorn P., van Steeg H.,
RA Grivell L.A., Slater E.C.;
RA "Nucleotide sequence analysis of the nuclear gene coding for
RT manganese superoxide dismutase of yeast mitochondria, a gene
RT previously assumed to code for the Rieske iron-sulphur protein.";
RL Eur. J. Biochem. 147:153-161(1985).
[2]
SEQUENCE FROM N.A.
RA STRAIN=S288C / AB972;
RA MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Faville A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
[3]
SEQUENCE OF 1-39 FROM N.A.
RA MEDLINE=89211942; PubMed=3072251;
RA Schrank I.S., Sims P.F., Oliver S.G.;
RA "Functional expression of the yeast Mn-superoxide dismutase gene in
RT Escherichia coli requires deletion of the signal peptide sequence.";
RL Gene 73:121-130(1988).
[4]
SEQUENCE OF 27-233.
RA Ditlow C., Johansen J.T., Martin B.M., Svendsen I.;
RA "The complete amino acid sequence of manganese-superoxide dismutase
RT from Saccharomyces cerevisiae.";
RL Carlsberg Res. Commun. 47:81-91(1982).
CC -!- FUNCTION: Destroys radicals which are normally produced within the

cells and which are toxic to biological systems.
-!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
-!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
-!- SUBUNIT: Homotetramer.
-!- SUBCELLULAR LOCATION: Mitochondrial matrix.
-!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X02156; CAA26092.1; -.
EMBL; U10400; AAB68939.1; -.
EMBL; M24079; AAA35065.1; -.
PIR; A00521; DGBYN.
HSP; P04179; IABM.
GeneOnline; 139325; -.
SGD; S0001050; SOD2.
DR GO; GO:0005759; C:mitochondrial matrix; IDA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IDA.
DR GO; GO:0006800; P:oxygen and reactive oxygen species metabolism; IMP.
DR InterPro: IPR001189; SODismutase.
DR Pfam: PF00081; sodfe; 1.
DR PRINTS: PF02777; sodfe; 1.
DR PRODOM: PD000475; SODismutase; 1.
DR PROSITE: PS00088; SOD MN; 1.
KW Oxidoreductase; Metal-binding; Manganese; Mitochondrion;
FT TRANSIT 1 25 MITOCHONDRION
FT CHAIN 27 233 SUPEROXIDE DISMUTASE [MN].
FT METAL 52 52 MANGANESE (BY SIMILARITY).
FT METAL 107 107 MANGANESE (BY SIMILARITY).
FT METAL 194 194 MANGANESE (BY SIMILARITY).
FT METAL 198 198 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 233 AA; 25774 MW; 88A9391FBB31D06E CRC64;

Query Match 2.9%; Score 7; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 39 ALEPYIS 45

RESULT 25
SODM_NEUCR
ID SODM_NEUCR STANDARD; PRT; 245 AA.
AC Q9Y7B3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
GN SOD-2 OR 18F11.030.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
[1]
[2]
SEQUENCE FROM N.A.
RA Dvorachek W.H., Natvig D.N.;
RA "Characterization of sod-2, the Neurospora crassa gene for manganese
RT superoxide dismutase.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA STRAIN=74-OR23-1A / FGSC 987;
RC
```

```

RX MEDLINE=22542210; PubMed=12555011;
RA Nannhauf G., Monrone C., Haase D., Mewes H.-W., Aign V.,
RA Heisele J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
RA Schulte U.;
RT "What's in the genome of a filamentous fungus? Analysis of the
RL Neurospora genome sequence.";
CC -1- NUCLEIC ACIDS Res. 31:1944-1954(2003).
CC -1- FUNCTION: Destroy radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF118809; AAD28503.1; -.
CC EMBL; AL670011; CAD21408.1; -.
CC HSSP; P04179; IABM.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodef; 1.
CC Pfam; PF02777; sodef; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC ProDom; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Mitochondrion;
KW TRANSIT PEPTIDE.
KW CHAIN 33 245 MITOCHONDRION (POTENTIAL).
KW METAL 58 58 SUPEROXIDE DISMUTASE [MN].
KW METAL 106 106 MANGANESE (BY SIMILARITY).
KW METAL 196 196 MANGANESE (BY SIMILARITY).
KW METAL 200 200 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 245 AA; 27019 MW; FF288947FB7676AD CRC64;

Query Match 2.9%; Score 7; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 45 ALEPYIS 51
|||||

RESULT 26
YACF SALTY STANDARD; PRT; 247 AA.
AC Q8XEA;
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Hypothetical protein YACF.
GN YACF OR STM0139 OR STY0161 OR T0145.
OS Salmonella typhimurium, and
CC Salmonella typhi.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Salmonella.
CC NCBI_TaxID=602, 601;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=1177609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RL Nature 413:852-856(2001).
RN [2]
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=1177608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RL enterica serovar Typhi CT18.";
RN Nature 413:848-852(2001).
RN [3]
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL and CT18.";
RN J. Bacteriol. 185:2330-2337(2003).
CC -1- SIMILARITY: Belongs to the UPF0289 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008700; AAL19103.1; -.
CC EMBL; AL627265; CAD01298.1; -.
CC EMBL; AE016834; AAO67877.1; -.
CC StyGene; SG????; yacF.
CC HAMAP; MF_01092; -; 1.
CC Hypothetical protein; Complete proteome.
SQ SEQUENCE 247 AA; 28425 MW; E1B9826AD004BE48 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212
Db 157 WLASLNP 163
|||||

RESULT 27
CFAD MOUSE STANDARD; PRT; 259 AA.
AC P03953; Q61280;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Complement factor D precursor (EC 3.4.21.46) (C3 convertase activator)
DE (Properdin factor D) (Adipsin) (28 kDa protein, adipocyte).
GN DF OR ADN.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=86278164; PubMed=3015943;
```


DR InterPro: IPR003000; SIR2.
 DR Pfam: PF02146; SIR2; 1.
 DR PROSITE: PS50305; SIRTUIN; 1.
 KW Hydroxylase; NAD; transcription regulation; Repressor; Nuclear protein;
 FT Metal-binding; Zinc.
 FT DOMAIN 83 370 DEACTYLASE SIRTUIN-TYPE.
 FT ACT SITE 213 213 BY SIMILARITY.
 FT METAL 221 221 ZINC (BY SIMILARITY).
 FT METAL 224 224 ZINC (BY SIMILARITY).
 FT METAL 251 251 ZINC (BY SIMILARITY).
 FT METAL 254 254 ZINC (BY SIMILARITY).
 SQ SEQUENCE 370 AA; 41765 MW; 86BB0238BFA914F1 CRC64;
 Query Match 2.9%; Score 7; DB 1; Length 370;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 KPIPTV 224
 DB 204 KPIPTV 210
 RESULT 29
 YG4S YEAST
 ID YG4S YEAST STANDARD; PRT; 409 AA.
 AC P5082;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical 46.7 kDa protein in PETS4-DIE2 intergenic region.
 GN YGR225W OR G841.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=98267763; PubMed=8701610;
 RA van der Aart O.J.M., Kleine K., Steensma H.Y.;
 RT "Sequence analysis of the 43 kb CRM1-YLM9-PET54-DIE2-SM11-PHO81-YHB4-
 RT PKI region from the right arm of Saccharomyces cerevisiae chromosome
 RT VII.";
 RL Yeast 12:385-390 (1996).
 CC -!- SIMILARITY: SOME, TO YEAST CDC20 AND S.POMBE SPAC1366.08.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X87941; CAA61174.1; -;
 CC DR EMBL; Z73010; CAA97253.1; -;
 CC DR PIR; S57689; S57689.
 CC DR GerMOnline; 141537; -;
 CC DR SGD; S0003457; AXA1.
 CC DR InterPro; IPR000002; Fizzy.
 CC DR Problem; PD004563; Fizzy; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 409 AA; 46660 MW; 91F7A246A28924D6 CRC64;
 Query Match 2.9%; Score 7; DB 1; Length 409;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 88 ASRNDYS 94
 DB 112 ASRNDYS 118

RESULT 30
 YDW1 SCHPO
 ID YDW1 SCHPO STANDARD; PRT; 441 AA.
 AC Q13909;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C23C11.01 in chromosome I.
 GN SPAC23C11.01.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voicikart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880 (2002).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: SOME, TO YEAST YIL090W.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z98559; CAB11154.1; -;
 CC DR PIR; T38239; T38239.
 CC DR GeneDB_Spombe; SPAC23C11.01; -;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 88 108 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT TRANSMEM 192 212 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 247 267 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 FT TRANSMEM 335 355 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 SQ SEQUENCE 441 AA; 49169 MW; 12B9B0095A93B3AA CRC64;
 Query Match 2.9%; Score 7; DB 1; Length 441;

Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TGOALAS 174
DB 430 TGOALAS 436

RESULT 31
AROA MYCTU
ID AROA MYCTU STANDARD; PRT; 450 AA.
AC P22457;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR RV3227 OR MT3324 OR MTCV20511.02.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=91072223; PubMed=2123856;
RA Garbe T., Jones C., Charles I.G., Dougan G., Young D.;
RT "Cloning and characterization of the *aroA* gene from *Mycobacterium*
tuberculosis.";
RL J. Bacteriol. 172:6774-6782(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=989395987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of *Mycobacterium tuberculosis* clinical and
laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
Phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
sixth step.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC

DR EMBL; X52269; CAA36510.1; -;
DR EMBL; M62708; AAB25356.1; -;
DR EMBL; Z95121; CAB0328.1; -;
DR EMBL; AE007144; AAK47667.1; -;
DR PIR; E70590; E70590.
DR TIGR; MT3324; -;
DR TubercuList; RV3227; -;
DR HAVAP; MF 00210; -; 1.
DR InterPro; IPR006264; AROA.
DR InterPro; IPR001986; EPSP_synth.
DR Pfam; PF00275; EPSP_synthase; 1.
DR PRODOM; PD001867; EPSP_synthase; 1.
DR TIGRFAMs; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP_SYNTHASE 1; 1.
DR PROSITE; PS00885; EPSP_SYNTHASE 2; 1.
KW Aromatic amino acid biosynthesis; transferase; Complete proteome.
SQ SEQUENCE 450 AA; 46425 MW; 27B866F9412A07D5 CRC64;
Query Match 2.9%; Score 7; DB 1; Length 450;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
DB 321 ALASPGS 327

RESULT 32
CA44 BOVIN
ID CA44 BOVIN STANDARD; PRT; 453 AA.
AC Q29422;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Lens;
RX MEDLINE=92112769; PubMed=1370461;
RA Mariyama M., Kalluri R., Hudson B.G., Readers S.T.;
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
collagens.";
RL J. Biol. Chem. 267:1253-1258(1992).
RN [2]
RP SEQUENCE OF 217-246.
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 217-233.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).

```
CC -1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77480; AAA30458.2; ALT_SEQ.
DR PIR; S18804; S18804.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 4.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SW00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1
FT DOMAIN <1 222 TRIPLE-HELICAL REGION.
FT DOMAIN 223 453 NONHELICAL REGION (NC1).
FT DISULFID 243 332 OR 329 (BY SIMILARITY).
FT DISULFID 276 329 OR 332 (BY SIMILARITY).
FT DISULFID 288 294 BY SIMILARITY.
FT DISULFID 351 449 OR 446 (BY SIMILARITY).
FT DISULFID 385 446 OR 449 (BY SIMILARITY).
FT DISULFID 397 404 BY SIMILARITY.
FT CONFLICT 219 219 I -> P (IN REF. 2 AND 3).
SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A659C1 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 453;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
DB 381 SPGSCLE 387
|||||

RESULT 33
GATB_RALSO
ID_GATB_RALSO STANDARD; PRT; 488 AA.
AC Q8Y3C6; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
DE (EC 6.3.5.-) (Asp/Glu-ADT subunit B).
GN GATB OR RSC0054 OR RS01877.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
```

```
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandin C., Choise N., Claudel-Renard C., Cunha S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002)
CC -1- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction
CC takes place in the presence of glutamine and ATP through an
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.
CC -1- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).
CC -1- SIMILARITY: Belongs to the gatB/gatE family. GatB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL646057; CAD13582.1; -.
DR HAWAP; MF_00121; -.
DR InterPro; IPR004413; GatB.
DR InterPro; IPR006107; GatB cent.
DR InterPro; IPR006075; GatB N.
DR InterPro; IPR003789; GatB_Yqey.
DR Pfam; PF01162; GatB; 1.
DR Pfam; PF02934; GatB_N; 1.
DR Pfam; PF02637; GatB_Yqey; 1.
DR TIGRFAMs; TIGR00133; GatB; 1.
DR PROSITE; PS01234; GATB; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 488 AA; 52342 MW; 657945DE961FA21B CRC64;

Query Match 2.9%; Score 7; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GELEKII 233
DB 429 GELEKII 435
|||||

RESULT 34
HEMA_IABAN
ID_HEMA_IABAN STANDARD; PRT; 550 AA.
AC P03441; Q83961; Q83962;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Bangkok/1/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11325;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82033259; PubMed=6169840;
RA Boch G.W., Sleight M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong
RT subtype influenza viruses during antigenic drift.";
```



```
DR HSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap.coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CARBOHYD 8
FT CARBOHYD 22
FT CARBOHYD 38
FT CARBOHYD 165
FT CARBOHYD 285
FT CARBOHYD 483
FT CONFLICT 137
FT SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 358 SEGTGOA 364

RESULT 37
HEMA_IADH3
ID HEMA_IADH3 STANDARD; PRT; 550 AA.
AC P12584; Q84012; Q89793;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11359;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16739; AAA43145.1; -.
DR HSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap.coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CARBOHYD 8
FT CARBOHYD 22
FT CARBOHYD 38
FT CARBOHYD 165
FT CARBOHYD 285
FT CARBOHYD 483
FT CONFLICT 137
FT SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 358 SEGTGOA 364

RESULT 38
HEMA_IADH4
ID HEMA_IADH4 STANDARD; PRT; 550 AA.
AC P12585; Q84013; Q84014;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16740; AAA43146.1; -.
DR HSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap.coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CARBOHYD 8
FT CARBOHYD 22
FT CARBOHYD 38
FT CARBOHYD 165
FT CARBOHYD 285
FT CARBOHYD 483
FT CONFLICT 137
FT SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;
```

```
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CARBOHYD 8
FT CARBOHYD 22
FT CARBOHYD 38
FT CARBOHYD 165
FT CARBOHYD 285
FT CARBOHYD 483
FT CONFLICT 137
FT SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 358 SEGTGOA 364

RESULT 39
HEMA_IADH4
ID HEMA_IADH4 STANDARD; PRT; 550 AA.
AC P12585; Q84013; Q84014;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16740; AAA43146.1; -.
DR HSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap.coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CARBOHYD 8
FT CARBOHYD 22
FT CARBOHYD 38
FT CARBOHYD 165
FT CARBOHYD 285
FT CARBOHYD 483
FT CONFLICT 137
FT SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;
```

```
SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 39
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11361;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawachi Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16741; AAA43147.1; -
DR PIR; E27813; HMIV21.
DR HSSP; P03437; 2VU.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328
FT CHAIN 330 550
FT CARBOHYD 7 7
FT CARBOHYD 8 8
FT CARBOHYD 22 22
FT CARBOHYD 38 38
FT CARBOHYD 165 165
FT CARBOHYD 285 285
FT CARBOHYD 483 483
FT CONFLICT 178 179
FT CONFLICT 388 388
FT K -> T (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15BF8C CRC64;
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 41
HEMA_IADH7 STANDARD; PRT; 550 AA.
ID HEMA_IADH7
```

```
QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 40
HEMA_IADH6 STANDARD; PRT; 550 AA.
ID HEMA_IADH6
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawachi Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16742; AAA43148.1; -
DR PIR; F27813; HMIV98.
DR HSSP; P03437; IHGJ.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328
FT CHAIN 330 550
FT CARBOHYD 22 22
FT CARBOHYD 38 38
FT CARBOHYD 165 165
FT CARBOHYD 285 285
FT CARBOHYD 483 483
FT CONFLICT 8 8
FT Y -> N (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 41
HEMA_IADH7 STANDARD; PRT; 550 AA.
ID HEMA_IADH7
```

AC P12588; Q84018; Q89470;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97265458; PubMed=2440178;
 RA Kida H., Kawakita Y., Naeve C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:1109-119(1987).
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
 CC cell receptors and for initiating infection.
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
 CC (HA1 and HA2) linked by a disulfide bond.
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M16743; AAA43149.1; -
 DR HSSP; P03437; 3HWG.
 DR InterPro; IPR008980; Capsid hemag.
 DR InterPro; IPR001364; Hemagglutn.
 DR InterPro; IPR008975; Viral_cap_coat.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutn; Glycoprotein.
 FT NON TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61761 MW; 65F81793281D53EB CRC64;
 Query Match 2.9%; Score 7; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 165 SEGTGQA 171
 DB 358 SEGTGQA 364
 RESULT 42
 HEMA_IADHK STANDARD; PRT; 550 AA.
 AC P43257;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11364;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 RT of avian H3 influenza viruses to pigs in southern China, where the
 RT A/Hong Kong/68 (H3N2) strain emerged.";
 RL J. Gen. Virol. 72:2007-2010(1991).
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
 CC cell receptors and for initiating infection.
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
 CC (HA1 and HA2) linked by a disulfide bond.
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D00929; BAA00769.1; -
 DR HSSP; P03437; 2VU.
 DR InterPro; IPR008980; Capsid hemag.
 DR InterPro; IPR001364; Hemagglutn.
 DR InterPro; IPR008975; Viral_cap_coat.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutn; Glycoprotein.
 FT NON TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61549 MW; 864639B829FE1B9A9 CRC64;
 Query Match 2.9%; Score 7; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 165 SEGTGQA 171
 DB 358 SEGTGQA 364
 RESULT 43
 HEMA_IADHL STANDARD; PRT; 550 AA.
 AC P43258;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hong Kong/64/76).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=45412;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 RT of avian H3 influenza viruses to pigs in southern China, where the

```

RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00932; BAA00771.1; --
DR HSSP; P03437; 2VU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; Hemagglutn12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61762 MW; 67EF8B49488C191A CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
|||||

RESULT 45
HEMA_IAGHK
ID HEMA_IAGHK STANDARD; PRT; 550 AA.
AC P43260;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=45411;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00930; BAA00770.1; --
DR HSSP; P03437; 2VU.

```



```
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral cap_coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD00225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 46
HEMA_IAMB6 STANDARD; PRT; 550 AA.
AC P12589;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Memphis/6/86).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11440;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306236; PubMed=3407150;
RA Katz J.M., Webster R.G.;
RT "Antigenic and structural characterization of multiple subpopulations
of H3N2 influenza virus from an individual.";
RL Virology 165:446-456(1988).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
(HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M21648; AAA43275.1; -
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD00225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 47
HEMA_IABZ2 STANDARD; PRT; 550 AA.
AC P11133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11497;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
in China.";
RL Virology 162:160-166(1988).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
(HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M19057; AAA43212.1; -
DR FIR; B29971; HMIVS3.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD00225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61804 MW; 52C9F14B309310ED CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364
```

```
SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 48
HEMA IAH3
ID HEMA IAH3 STANDARD; PRT; 550 AA.
AC P1134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.";
RL Virology 162:160-166(1988).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19056; AAA43211.1; ALT_TERM.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TERM 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61580 MW; 991F6D8BC02F24F2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 49
HEMA IAHAL
ID HEMA IAHAL STANDARD; PRT; 565 AA.
AC P16994; Q83990; Q83991;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
GN HA
OS Influenza A virus (strain A/Equine/Algiers/72).
OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 163:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24721; AAA43100.1; ALT_SEQ.
DR HSSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 340 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63831 MW; BA533050DC3F186B CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 50
HEMA IAHFO
ID HEMA IAHFO STANDARD; PRT; 565 AA.
AC P16995; Q83992; Q83993;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
```

```
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Fontainebleau/76) (Influenza A
OS virus (strain A/Equine/France/1/76)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11399;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.";
RL J. Virol. 66:1129-1138(1992).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24723; AAA43101.1; AUT_SEQ.
DR EMBL; M73773; -; NOT_ANNOTATED_CDS.
DR HSSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 13 15 WVY -> AVD (IN REF. 2).
FT CONFLICT 20 20 T -> I (IN REF. 2).
FT CONFLICT 150 150 R -> G (IN REF. 2).
FT CONFLICT 187 187 N -> D (IN REF. 2).
FT CONFLICT 242 242 S -> A (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 479 479 N -> G (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63686 MW; 1BB06B765992E87C CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEGTOGA 171
Db 373 SEGTOGA 379

RESULT 52
HEMA IAHK6 STANDARD; PRT; 565 AA.
AC P16959;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Kentucky/2/86).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11403;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24727; -; NOT_ANNOTATED_CDS.
DR HSSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 13 15 WVY -> AVD (IN REF. 2).
FT CONFLICT 20 20 T -> I (IN REF. 2).
FT CONFLICT 150 150 R -> G (IN REF. 2).
FT CONFLICT 187 187 N -> D (IN REF. 2).
FT CONFLICT 242 242 S -> A (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 479 479 N -> G (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63610 MW; 2038CC1C6C9B88C5 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEGTOGA 171
Db 373 SEGTOGA 379

RESULT 52
HEMA IAHK7 STANDARD; PRT; 565 AA.
AC P16956; Q83994; Q83995;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
```

```

GN HA.
OS Influenza A virus (strain A/Equine/Kentucky/1/87).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24722; AAA43107.1; ALT_SEQ.
DR HSSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63702 MW; 93963AF456486787 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379
|||||

RESULT 53
HEMA_IAHNM STANDARD; PRT; 565 AA.
AC P16997; Q83996; Q83997;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Romania/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24728; AAA43103.1; ALT_SEQ.
DR HSSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63702 MW; 93963AF456486787 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379
|||||

RESULT 54
HEMA_IAHRO STANDARD; PRT; 565 AA.
AC P16998; Q83998; Q83999;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Romania/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M24724; AAA43109.1; ALT_SEQ.
 DR HSSP; P03437; 1HTM.
 DR InterPro; IPR008980; Capsid hemag.
 DR Pfam; PF000329; Hemagglutn.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 565 AA; 63660 MW; 91A07D33FAFDC842 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 SEGTGQA 171
 DB 373 SEGTGQA 379

RESULT 55

HEMA_IAHSU
 ID HEMA_IAHSU STANDARD; PRT; 565 AA.
 AC P16999; Q84000; Q84001;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Equine/Santiago/1/85).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11414;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=9204899; PubMed=2705299;
 RA Kawaoka Y., Bean W.J., Webster R.G.;
 RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
 RL Virology 169:283-292 (1989).
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
 CC cell receptors and for initiating infection.
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
 CC (HA1 and HA2) linked by a disulfide bond.
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M24725; AAA43110.1; ALT_SEQ.
 DR HSSP; P03437; 1HTM.
 DR InterPro; IPR008980; Capsid hemag.
 DR Pfam; PF000329; Hemagglutn.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN (BY SIMILARITY).
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN (BY SIMILARITY).
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).

DR Pfam; PF000329; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 565 AA; 63665 MW; 399FABF4BA231327 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 SEGTGQA 171
 DB 373 SEGTGQA 379

RESULT 56

HEMA_IAHSU
 ID HEMA_IAHSU STANDARD; PRT; 565 AA.
 AC Q08011;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Equine/Suffolk/89).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=45413;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93277383; PubMed=8503788;
 RA Birns M.M., Daly J.M., Charnside E.D., Mumford J.A., Wood J.M.,
 RA Richards C.M., Daniels R.S.;
 RT "Genetic and antigenic analysis of an equine influenza H 3 isolate
 RT from the 1989 epidemic.";
 RL Arch. Virol. 130:33-44 (1993).
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
 CC cell receptors and for initiating infection.
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
 CC (HA1 and HA2) linked by a disulfide bond.
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X58437; CAA48482.1; -.
 DR HSSP; P03437; 1HTM.
 DR InterPro; IPR008980; Capsid hemag.
 DR Pfam; PF000329; Hemagglutn.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN (BY SIMILARITY).
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN (BY SIMILARITY).
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63699 MW; C7A4E3B54B87D1A1 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 57
HEMA IAHTO
ID HEMA IAHTO STANDARD; PRT; 565 AA.
AC P17001; Q84004; Q84005;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Tennessee/5/86)
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11417;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC HSSP; P03437; IHTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63698 MW; 1FB4485F0E7AC2C4 CRC64;
```

```
Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 58
HEMA IAHTO
ID HEMA IAHTO STANDARD; PRT; 565 AA.
AC P17000; Q84002; Q84003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11418;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24720; AAA43111.1; ALT_SEQ.
CC HSSP; P03437; IHTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63580 MW; 84BD74AD70629B7A CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 373 SEGTGQA 379
```

RESULT 59

HEMA_IAHUR STANDARD; PRT; 565 AA.
AC P17002; Q84006; Q84007;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain]
GN HA.
OS Influenza A virus (strain A/Equine/Uruguay/1/63).
OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11419;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=27052299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; M24718; AAA43114.1; ALT_SEQ.
DR HSP; P03437; IHTW.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL
FT CHAIN 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63604 MW; 32818573564C8F94 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171

Db 373 SEGTGQA 379

RESULT 60

HEMA_IAAIC STANDARD; PRT; 566 AA.
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain]
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).

OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
RA Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
RT influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
RL Nature 286:771-776(1980).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=81123029; PubMed=7464906;
RA Wilson I.A., Skehel J.J., Wiley D.C.;
RT "Structure of the haemagglutinin membrane glycoprotein of influenza
RT virus at 3-A resolution.";
RL Nature 289:366-373(1981).
RN [3]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=88232903; PubMed=3374584;
RA Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA Wiley D.C.;
RT "Structure of the influenza virus haemagglutinin complexed with its
RT receptor, sialic acid.";
RL Nature 333:426-431(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX MEDLINE=90107940; PubMed=2295311;
RA Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA Wiley D.C.;
RT "The structure of a membrane fusion mutant of the influenza virus
RT haemagglutinin.";
RL EMBO J. 9:17-24(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=90230310; PubMed=2329580;
RA Weis W.I., Bruegger A.T., Skehel J.J., Wiley D.C.;
RT "Refinement of the influenza virus hemagglutinin by simulated
RT annealing.";
RL J. Mol. Biol. 212:737-761(1990).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=94352388; PubMed=8072525;
RA Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT "Structure of influenza haemagglutinin at the pH of membrane fusion.";
RL Nature 371:37-43(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
RX MEDLINE=98120975; PubMed=9461077;
RA Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT "Antigen distortion allows influenza virus to escape neutralization.";
RL Nat. Struct. Biol. 5:119-123(1998).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; J02090; AAA43178.1; --
DR EMBL; V01085; CAA24269.1; --
DR PIR; A33231; HMTVHA.
DR PDB; 2HMG; 31-OCT-93.
DR PDB; 3HMG; 31-OCT-93.
DR PDB; 4HMG; 31-OCT-93.

DR PDB; SHMG; 31-JAN-94.
 DR PDB; IHGD; 31-JAN-94.
 DR PDB; IHGE; 31-JAN-94.
 DR PDB; IHGF; 31-JAN-94.
 DR PDB; IHGG; 31-JAN-94.
 DR PDB; IHGH; 31-JAN-94.
 DR PDB; IHGI; 31-JAN-94.
 DR PDB; IHGJ; 31-JAN-94.
 DR PDB; IHM; 14-FEB-95.
 DR PDB; IHM; 29-APR-98.
 DR PDB; 2VIS; 29-APR-98.
 DR PDB; 2VIT; 29-APR-98.
 DR PDB; 2VIU; 29-APR-98.
 DR PDB; 1E08; 29-NOV-00.
 DR PDB; 1HA0; 22-DEC-99.
 DR PDB; 1J8H; 13-MAR-02.
 DR PDB; 1KEN; 24-APR-02.
 DR PDB; 1QFU; 29-DEC-99.
 DR PDB; 1QUL; 05-JAN-00.
 DR InterPro; IPR008980; Capsid hemag.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin: 1.
 DR PRINTS; PR00329; HEMAGGLUTN1.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 16
 FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
 FT DISULFID 30 482 INTERCHAIN.
 FT DISULFID 68 293
 FT DISULFID 80 92
 FT DISULFID 113 155
 FT DISULFID 297 321
 FT DISULFID 489 493
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .).
 FT STRAND 27 35
 FT STRAND 40 42
 FT STRAND 50 52
 FT STRAND 55 57
 FT STRAND 59 60
 FT STRAND 67 70
 FT STRAND 74 76
 FT TURN 78 79
 FT TURN 82 87
 FT TURN 88 88
 FT HELIX 90 95
 FT TURN 96 97
 FT TURN 99 99
 FT STRAND 102 105
 FT TURN 107 108
 FT STRAND 116 117
 FT TURN 119 120
 FT HELIX 121 131
 FT STRAND 133 133
 FT STRAND 136 138
 FT TURN 144 145
 FT STRAND 146 147
 FT STRAND 152 157
 FT TURN 158 159
 FT STRAND 160 162
 FT TURN 165 166
 FT STRAND 167 169
 FT STRAND 171 173
 FT TURN 174 175
 FT STRAND 176 176
 FT STRAND 180 185
 FT STRAND 192 200

FT HELIX 204 211
 FT STRAND 217 221
 FT STRAND 226 229
 FT STRAND 239 239
 FT TURN 240 241
 FT STRAND 242 242
 FT STRAND 245 253
 FT TURN 255 256
 FT STRAND 258 265
 FT STRAND 267 270
 FT STRAND 272 275
 FT STRAND 282 285
 FT STRAND 290 294
 FT STRAND 297 299
 FT TURN 300 301
 FT STRAND 302 304
 FT STRAND 310 311
 FT STRAND 318 320
 FT STRAND 323 324
 FT STRAND 331 333
 FT STRAND 337 337
 FT TURN 347 348
 FT TURN 350 350
 FT STRAND 351 351
 FT TURN 352 354
 FT STRAND 355 355
 FT STRAND 359 360
 FT TURN 361 361
 FT STRAND 367 373
 FT TURN 374 375
 FT STRAND 376 382
 FT HELIX 383 400
 FT TURN 401 401
 FT STRAND 406 407
 FT HELIX 421 471
 FT HELIX 472 474
 FT STRAND 475 477
 FT STRAND 482 485
 FT HELIX 491 498
 FT TURN 499 500
 FT HELIX 504 515
 FT TURN 518 519
 SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;
 Query Match 2.9%; Score 7; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred.No.39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 165 SEGTGQA 171
 Db 374 SEGTGQA 380
 RESULT 61
 HEMA_IADA3
 ID_HEMA_IADA3 STANDARD; PRT; 566 AA.
 AC P26134;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA
 OS Influenza A virus (strain A/Duck/Alberta/78/76).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11348;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92114135; PubMed=1731092;
 RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
 RA Webster R.G.;
 RT "Evolution of the H3 influenza virus hemagglutinin from human and


```

RT nonhuman hosts.":
RL J. Virol. 66:1129-1138 (1992).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; M73771; -; NOT_ANNOTATED_CDS.
CC HSSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC FT SIGNAL 1 16
CC FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
CC FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FF9415B89 CRC64;
CC
CC Query Match 2.9%; Score 7; DB 1; Length 566;
CC Best Local Similarity 100.0%; Pred. No. 39;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 165 SEGTGQA 171
CC |||||
CC Db 374 SEGTGQA 380
CC
CC RESULT 62
CC HEMA_IADM2
CC ID HEMA_IADM2 STANDARD; PRT; 566 AA.
CC AC P26135;
CC DT 01-MAY-1992 (Rel. 22, Created)
CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
CC Hemagglutinin HA2 chain].
CC GN HA.
CC OS Influenza A virus (strain A/Duck/Memphis/328/74).
CC OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
CC OC Influenza A viruses; Influenzavirus A.
CC OX NCBI_TaxID=11367;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=92114135; PubMed=1731092;
CC RA Bean W.J., Schell M., Katz J., Kawaka Y., Naeve C., Gorman O.,
CC Webster R.G.;
CC RT "Evolution of the H3 influenza virus hemagglutinin from human and
CC nonhuman hosts."
CC RL J. Virol. 66:1129-1138 (1992).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; M73772; -; NOT_ANNOTATED_CDS.
CC HSSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC FT SIGNAL 1 16
CC FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
CC FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 566 AA; 63572 MW; C85DFC5DEB8D0D CRC64;
CC
CC Query Match 2.9%; Score 7; DB 1; Length 566;
CC Best Local Similarity 100.0%; Pred. No. 39;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 165 SEGTGQA 171
CC |||||
CC Db 374 SEGTGQA 380
CC
CC RESULT 63
CC HEMA_IADU3
CC ID HEMA_IADU3 STANDARD; PRT; 566 AA.
CC AC P03442;
CC DT 21-JUL-1986 (Rel. 01, Created)
CC DT 21-JUL-1986 (Rel. 01, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
CC Hemagglutinin HA2 chain].
CC GN HA.
CC OS Influenza A virus (strain A/Duck/Ukraine/1/63).
CC OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
CC OC Influenza A viruses; Influenzavirus A.
CC OX NCBI_TaxID=11374;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=82025542; PubMed=6169439;
CC RA Fang R., Min Jou W., Huybrecock D., Devos R., Fiers W.;
CC RT "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
CC gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
CC hemagglutinin."
CC RL Cell 25:315-323 (1981).
CC CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)

```

```
-----
EMBL; V01087; CAA24271.1; --
PDB; 11BN; 08-AUG-01.
PDB; 11BO; 08-AUG-01.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00509; Hemagglutinin; I.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 374 SEGTGOA 380
|||||

RESULT 64
HEMA IAEN7
ID HEMA IAEN7 STANDARD; PRT; 566 AA.
AC P03440;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/England/321/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11378;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83110955; PubMed=6822815;
RA Hauptmann R., Clarke L.D., Mountford R.C., Bachmayer H., Almond J.W.;
RT "Nucleotide sequence of the haemagglutinin gene of influenza virus
RT A/England/321/77.";
RL J. Gen. Virol. 64:215-220 (1983).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X05907; CAA29337.1; --
PDB; B92790; HMTV6.
PDB; B92790; HMTV6.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
-----
```

```
PRINTS; PR00329; HEMAGGLUTN12.
ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63608 MW; FA5B866FF4B8C888 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 374 SEGTGOA 380
|||||

RESULT 65
HEMA IAMAO
ID HEMA IAMAO STANDARD; PRT; 566 AA.
AC P26138;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Mallard/New York/6874/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11436;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.";
RL J. Virol. 66:1129-1138 (1992).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M73776; -- NOT_ANNOTATED_CDS.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
-----
```

```
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63574 MW; BB206011COBD9A3B CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 374 SEGTGQA 380

RESULT 66
HEMA_IAME1 STANDARD; PRT; 566 AA.
AC P03439;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Memphis/1/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11438;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83303827; PubMed=6193634;
RA Newton S.E., Air G.M., Webster R.G., Laver W.G.;
RT "Sequence of the hemagglutinin gene of influenza virus A/Memphis/1/71
RT and previously uncharacterized monoclonal antibody-derived
RT variants.";
RL Virology 128:495-501 (1983).
RN [2]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=82150925; PubMed=6174976;
RA Air G.M.;
RT "Sequence relationships among the hemagglutinin genes of 12 subtypes
RT of influenza A virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:7639-7643 (1981).
CC -1- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -1- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -1- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J02132; AAA43187.1; -.
CC HSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTIN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63264 MW; 1D9313AB3C380CD7 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
```

```
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63558 MW; 6383C6F9D81AB941 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 374 SEGTGQA 380

RESULT 67
HEMA_IAME2 STANDARD; PRT; 566 AA.
AC P03439;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Memphis/102/72).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11439;
RN [1]
RP SEQUENCE FROM N.A.
RA Sleight M.J., Both G.W., Brownlee G.G., Bender V.J., Moss B.A.;
RT "The haemagglutinin gene of influenza A virus: nucleotide sequence
RT analysis of cloned DNA copies.";
RL (In) Laver G., Air G. (eds.);
RL Structure and variation in influenza virus, pp.69-79, Elsevier,
RL New York (1980).
CC -1- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -1- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -1- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01089; CAA24273.1; -.
CC F1R; A94441; HMTVHM.
CC HSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTIN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63264 MW; 1D9313AB3C380CD7 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
```


Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTCQA 171
DB 374 SEGTCQA 380

RESULT 70

HEMA IAZUK STANDARD; PRT; 566 AA.
AC P26141;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Swine/Ukkel/1/84).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11517;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts";
RL J. Virol. 66:1129-1138 (1992).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M73775; -; NOT ANNOTATED_CDS.
CC HSSP; P03437; 1HTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; Hemagglutn12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 28 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 566 AA; 63725 MW; 44661EDB8B3D5B331 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTCQA 171
DB 374 SEGTCQA 380

RESULT 71

HEMA IAVI7 STANDARD; PRT; 567 AA.
AC P03435;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Victoria/3/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11483;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=80155186; PubMed=6153930;
RA Min Jou W., Verhoeven M., Devos R., Saman E., Fang R.,
RA Huybrecock D., Fiers W., Threlfall G., Barber C., Carey N.,
RA Entage S.;
RT "Complete structure of the hemagglutinin gene from the human
RT Influenza A/Victoria/3/75 (H3N2) strain as determined from cloned
RT DNA";
RL Cell 19:683-696 (1980).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Jou W.M., Devos R., Huybrecock D., Saman E.,
RA Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
RT influenza strains A/Aichi/2/68 and A/Victoria/3/75";
RL Nature 286:771-776 (1980).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01098; CAA24281.1; -
CC EMBL; V01085; CAA24270.1; -
CC PIR; A90794; HMIV.
CC HSSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; Hemagglutn12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 345 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 347 567 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 25 25 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 39 39 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 55 55 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 567 AA; 63422 MW; 824D98A880EC5DEF CRC64;

Query Match 2.9%; Score 7; DB 1; Length 567;

Best Local Similarity 100.0%; Pred. No. 39; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||
Db 375 SEGTGQA 381

RESULT 72
CA44 RABIT
ID CA44 RABIT STANDARD; PRT; 623 AA.
AC P55787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RX TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01477; -; NOT_ANNOTATED_CDS.
CC
CC PIR; A45137; A45137.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagen4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 5.
CC ProDom; PD003923; ProcollagenC4; 1.
CC SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON TER 1 1
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT DOMAIN 393 623 NONHELICAL REGION (NC1).
FT DISULFID 413 502 OR 499 (BY SIMILARITY).

FT DISULFID 446 499 OR 502 (BY SIMILARITY).
FT DISULFID 458 484 BY SIMILARITY.
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 OR 619 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SQ SEQUENCE 623 AA; 62393 MW; CCBC9BB31242FE82 CRC64;
Query Match 2.9%; Score 7; DB 1; Length 623;
Best Local Similarity 100.0%; Pred. No. 42; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
|||||
Db 551 SPGSCLE 557

RESULT 73
TOXA PSEAE
ID TOXA PSEAE STANDARD; PRT; 638 AA.
AC P11439; Q91417;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exotoxin A precursor (NAD-dependent ADP-ribosyltransferase)
DE (EC 2.4.2.-).
GN ETA OR PA1148.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 26-53.
RX MEDLINE=84194063; PubMed=6201861;
RA Gray G.L., Smith D.H., Baldrige J.S., Harkins R.N., Vasil M.L.,
RA Chen E.Y., Heyneker H.L.;
RT "Cloning, nucleotide sequence, and expression in *Escherichia coli* of
RT the exotoxin A structural gene of *Pseudomonas aeruginosa*.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:2645-2649(1984).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [3]
ACTIVE SITE.
RX MEDLINE=87250491; PubMed=2885323;
RA Carroll S.F., Collier R.J.;
RT "Active site of *Pseudomonas aeruginosa* exotoxin A. Glutamic acid 553
RT is photolabeled by NAD and shows functional homology with glutamic
RT acid 148 of diphtheria toxin.";
RN J. Biol. Chem. 262:8707-8711(1987).
RN [4]
DOMAINS.
RX MEDLINE=90375493; PubMed=2118903;
RA Chaudhary V.K., Jinno Y., Galo M.G., Fitzgerald D., Pastan I.;
RT "Mutagenesis of *Pseudomonas aeruginosa* exotoxin in identification of sequences
RT responsible for the animal toxicity.";
RN J. Biol. Chem. 265:16306-16310(1990).
RN [5]
DOMAINS.
RX MEDLINE=91006124; PubMed=2170123;
RA Bourdenet S., Vacheron M.-J., Guinand M., Michel G., Arminjon F.;
RT "Biochemical and immunochemical studies of proteolytic fragments of
RT exotoxin A from *Pseudomonas aeruginosa*.";
RL Eur. J. Biochem. 192:379-385(1990).


```

[3]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP TISSUE=Eye, and Lymph;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Bhat N.K.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]
SEQUENCE OF 1-354 FROM N.A. (ISOFORM 3).
RC TISSUE=Carcinoma;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
CC -!- FUNCTION: Catalyzes the phosphorylation of sphingosine to form
CC sphingosine 1-phosphate (SPP), a lipid mediator with both intra-
CC and extracellular functions. Also acts on D-erythro-
CC dihydro-sphingosine, D-erythro-sphingosine and L-threo-
CC dihydro-sphingosine.
CC -!- CATALYTIC ACTIVITY: Sphingosine + ATP = sphingosine 1-phosphate +
CC ADP.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC IsoId=Q9NRA0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NRA0-2; Sequence=VSP_006217;
CC Name=3;
CC IsoId=Q9NRA0-3; Sequence=VSP_006218;
CC -!- SIMILARITY: Contains 1 DAGKc domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF245447; AAF74124.1; -
CC EMBL; AL136701; CAB66636.1; -
CC EMBL; BC006161; ARI06161.1; -
CC EMBL; BC010671; ARI0671.1; -
CC EMBL; AK000599; BAA91280.1; -
CC Genew; HGNC:18859; SPHK2.
CC MIM; 607092; -
CC GO; GO:0005829; C.cytosol; IEP.
CC GO; GO:0005624; C.membrane fraction; IEP.
CC GO; GO:0008189; F.apoptosis inhibitor activity; NAS.
CC GO; GO:0005515; F.protein binding; IPI.
CC GO; GO:0017016; F.Ras interactor activity; NAS.
CC GO; GO:0008481; F.sphinganine kinase activity; NAS.
CC GO; GO:0006916; P.anti-apoptosis; NAS.

```

```

DR GO; GO:0008283; P.cell proliferation; NAS.
DR GO; GO:0006869; P.sphinganine-1-phosphate biosynthesis; NAS.
DR InterPro; IPR001206; DAGKc.
DR Pfam; PF00781; DAGKc; 1.
DR ProDom; PD005043; DAGKc; 1.
DR SMART; SM00046; DAGKc; 1.
DR Transferase; Kinase; ATP-binding; Alternative splicing.
FT VARSPLIC 1 36 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 292 390 /FTID=VSP_006217.
FT VARSPLIC 292 390 /FTID=VSP_006218.
FT CONFLICT 49 49 P -> S (IN REF. 2).
SQ SEQUENCE 654 AA; 69217 MW; F73FFEC930DA50F CRC64;
Query Match 2.9%; Score 7; DB 1; Length 654;
Best Local Similarity 100.0%; Pred. No. 44; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
Qy 171 ALASPGS 177
Db 431 ALASPGS 437
RESULT 75
ID IGAA YERPE STANDARD; PRT; 715 AA.
AC P88722;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative membrane protein igaa homolog.
GN YPO0142 OR Y3922.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE IGAA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


or send an email to license@isb-sib.ch).

```
CC CC
DR EMBL; X81053; CAA56943.1; -.
DR EMBL; AB008496; BAA25065.1; -.
DR EMBL; D17391; BAA04214.1; -.
DR PIR; A55360; CGHUIB.
DR Genew; HGNC:2206; COL4A4.
DR MIM; 120131; -.
DR MIM; 143200; -.
DR MIM; 203780; -.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38
FT CHAIN 39 1690
FT DOMAIN 39 64
FT DOMAIN 65 1459
FT DOMAIN 1460 1690
FT SITE 94 96
FT SITE 145 147
FT SITE 189 191
FT SITE 310 312
FT SITE 724 726
FT SITE 785 787
FT SITE 989 991
FT SITE 1206 1207
FT SITE 1212 1214
FT SITE 1480 1569
FT DISULFID 1513 1566
FT DISULFID 1525 1531
FT DISULFID 1588 1686
FT DISULFID 1622 1683
FT DISULFID 1634 1641
FT CARBOHYD 142 144
FT CARBOHYD 669 669
FT VARIANT 441 446
FT VARIANT 545 545
FT VARIANT 570 570
FT VARIANT 897 897
FT VARIANT 931 931
FT VARIANT 1004 1004
FT VARIANT 1030 1030
FT VARIANT 1201 1201
FT VARIANT 1402 1402
FT VARIANT 1572 1572
FT CONFLICT 1659 1660
FT SEQUENCE 1690 AA; 164095 MW; ELE72F283A72BAE CRC64;
Query Match 2.9%; Score 7; DB 1; Length 1690;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 174 SPGSCLE 180
|||||
```

Db 1618 SPGSCLE 1624

RESULT 77

```
CA24 CAEEL STANDARD; PRT; 1758 AA.
ID P17140; Q19098; Q19099;
AC 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (lethal protein 2).
GN LET-2 OR CLB-1 OR FO1G12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RL J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guc X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RA Wu X., Le T.T.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANTS.
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RL EMBO J. 13:3278-3285(1994).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC Vital for embryonic development.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NCI
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I; Synonyms=a;
CC IsoId=P17140-1; Sequence=Displayed;
CC Name=II; Synonyms=b;
CC IsoId=P17140-2; Sequence=VSP_001160;
CC -!- DEVELOPMENTAL STAGE: Isoform I is predominant in embryos and
CC isoform II is predominant in the larvae and adults.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```



```
FT VARSPLIC 230 266 GCGGPGGPGGPPVSTGAKTTLIGPEGABGMKGEK -->
FT GCGGAGGPGGPPGPGPREFTGSGSIVGPRGSHGDKGVK (in
FT isoform II).
FT /FTIG-VSP_001159.
SQ SEQUENCE 1763 AA; 168526 MW; 304F528BC06A80D CRC64;

Query Match 2.9%; Score 7; DB 1; Length 1763;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 73 TWPFLFC 79
Db 1587 TWPFLFC 1593

RESULT 79
ID DMD_CHICK STANDARD; PRT; 3660 AA.
AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dystrophin.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal
coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the
plasma membrane.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 22 spectrin repeats.
CC -!- SIMILARITY: Contains 1 WW domain.
CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13369; CAA31746.1; -.
CC F01; S02041; S02041.
CC HSSP; P46939; LBHD.
CC InterPro; IPR001589; Actbind actnin.
CC InterPro; IPR001715; Calponin-like.
CC InterPro; IPR002017; Spectrin.
CC InterPro; IPR001202; WW_Rsp5_WWP.
CC InterPro; IPR000433; ZnF_ZZ.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00435; spectrin; 21.
CC Pfam; PF00397; WW; 1.
CC Pfam; PF00569; ZZ; 1.
```

```
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 21.
DR SMART; SM00456; WW; 1.
DR SMART; SM00291; ZnF_ZZ; 1.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00020; ACTININ 2; 1.
DR PROSITE; PS00021; CH; 2; 1.
DR PROSITE; PS01159; WW_DOMAIN 1; 1.
DR PROSITE; PS00020; WW_DOMAIN 2; 1.
DR PROSITE; PS01357; ZF_ZZ 1; 1.
DR PROSITE; PS0135; ZF_ZZ 2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
Repeat; Zinc-finger.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 138 241 CH 1.
FT REPEAT 341 449 SPECTRIN 1.
FT REPEAT 450 558 SPECTRIN 2.
FT REPEAT 561 669 SPECTRIN 3.
FT REPEAT 721 830 SPECTRIN 4.
FT REPEAT 832 936 SPECTRIN 5.
FT REPEAT 945 1047 SPECTRIN 6.
FT REPEAT 1050 1156 SPECTRIN 7.
FT REPEAT 1159 1265 SPECTRIN 8.
FT REPEAT 1268 1369 SPECTRIN 9.
FT REPEAT 1470 1570 SPECTRIN 10.
FT REPEAT 1573 1678 SPECTRIN 11.
FT REPEAT 1681 1782 SPECTRIN 12.
FT REPEAT 1879 1981 SPECTRIN 13.
FT REPEAT 2013 2103 SPECTRIN 14.
FT REPEAT 2214 2321 SPECTRIN 15.
FT REPEAT 2472 2574 SPECTRIN 16.
FT REPEAT 2577 2683 SPECTRIN 17.
FT REPEAT 2686 2799 SPECTRIN 18.
FT REPEAT 2802 2904 SPECTRIN 19.
FT REPEAT 2906 2928 SPECTRIN 20.
FT REPEAT 2931 3037 SPECTRIN 21.
FT REPEAT 3052 3085 SPECTRIN 22.
FT ZN_FING 3304 3351 ZZ-TYPE.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

Query Match 2.9%; Score 7; DB 1; Length 3660;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 208 ASLNPER 214
Db 2284 ASLNPER 2290

RESULT 80
ID CHLY_CARPA STANDARD; PRT; 25 AA.
AC P81241;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bifunctional chitinase/lysozyme [includes: Chitinase (EC 3.2.1.14);
DE Lysozyme (EC 3.2.1.17)] (Fragments).
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE OF 1-22.
RC TISSUE=latex;
RX MEDLINE=20063791; PubMed=10594185;
RA Subroto T., Sufiati S., Beintema J.J.;
```

RT "Papaya (Carica papaya) lysozyme is a member of the family 19 (basic,
RT class II) chitinases.";
RL J. Mol. Evol. 49:819-821 (1999).
RN [2]
RP SEQUENCE OF 1-5 AND 23-25.
RC TISSUE=Leaf;
RX MEDLINE=69130229; PubMed=5773045;
RA Howard J.B., Glazer A.N.;
RT "Papaya lysozyme. Terminal sequences and enzymatic properties.";
RL J. Biol. Chem. 244:1399-1409 (1969).
CC -!- FUNCTION: Bifunctional enzyme with lysozyme/chitinase activity.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to family 19 of glycosyl hydrolases.
DR InterPro: IPR000726; Glyco_hydro_19
DR PROSITE: PS00773; CHITINASE_19_1; PARTIAL.
DR PROSITE: PS00774; CHITINASE_19_2; PARTIAL.
KW Hydrolase; Glycosidase; Chitin degradation; Multifunctional enzyme.
FT CONFLICT 3 3 E -> S (IN REF. 2).
FT NON_CONS 22 23
FT NON_TER 25 25
SQ SEQUENCE 25 AA; 2877 MW; 33BA3F018F33ACD6 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EKILSR 235
Db 3 EKILSR 8

RESULT 81
PSAI CHAGL
ID PSAL CHAGL STANDARD; PRT; 36 AA.
AC Q8MX55;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit VIII (PSI-I).
GN PSAL
OS Chaetosphaeeridium globosum.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Coleochaetales;
OC Chaetosphaeeridiaceae; Chaetosphaeeridium.
OX NCBI_TaxID=96477;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M311;
RX MEDLINE=22177139; PubMed=12161560;
RA Turmel M., Otis C., Lemieux C.;
RT "The chloroplast and mitochondrial genome sequences of the charophyte
RT Chaetosphaeeridium globosum: insights into the timing of the events
RT that restructured organelle DNAs within the green algal lineage that
RT led to land plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:11275-11280 (2002).
CC -!- FUNCTION: May help in the organization of the psal subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the psal family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC

DR EMBL; AF494278; AAM96535.1; -.
DR HAMAP; MF_00431; -; 1.
DR InterPro; IPR01302; PSI_8.
DR Pfam; PF00796; PSI_8; 1.
KW Photosystem I; Photosynthesis; Transmembrane; Chloroplast.
FT TRANSMEM 5
FT POTENTIAL 27
SQ SEQUENCE 36 AA; 3876 MW; 963F419C07BD405D CRC64;

Query Match 2.5%; Score 6; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIA 133
Db 19 PAIAIA 24

RESULT 82
LA89 LACAC
ID LA89 LACAC STANDARD; PRT; 46 AA.
AC Q48501; Q9RSH2; -
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bacteriocin acidocin 8912 precursor.
GN ACDT.
OS Lactobacillus acidophilus.
OG Plasmid pLA103.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK8912;
RX MEDLINE=96140009; PubMed=8554765;
RA Kanatani K., Tahara T., Oshimura M., Sano K., Umezawa C.;
RT "Cloning and nucleotide sequence of the gene for acidocin 8912, a
RT bacteriocin from Lactobacillus acidophilus TK8912.";
RL Lett. Appl. Microbiol. 21:384-386 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TK8912;
RX MEDLINE=93005074; PubMed=1368836;
RA Tahara T., Kanatani K., Yoshida K., Miura H., Sakamoto M.,
RA Oshimura M.;
RT "Purification and some properties of acidocin 8912, a novel
RT bacteriocin produced by Lactobacillus acidophilus TK8912.";
RL Biosci. Biotechnol. Biochem. 56:1212-1215 (1992).
CC -!- FUNCTION: Has a bactericidal effect on sensitive cells but not a
CC bacteriolytic effect.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D43626; BAA07737.1; -.
DR EMBL; AB081463; BAB86322.1; -.
KW Antibiotic; Bacteriocin; Plasmid.
FT PROPEP 1 20
FT CHAIN 21 46 BACTERIOCIN ACIDOCIN 8912.
FT CONFLICT 33 33 W -> R (IN REF. 3).
FT CONFLICT 33 33
SQ SEQUENCE 46 AA; 5331 MW; EAF910D04D2AC3B8 CRC64;

CC

```
Query Match          2.5%; Score 6; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SLWKGK 155
Db 31 SLWKGK 36

RESULT 83
PSBZ_CYPAP
ID PSBZ_CYPAP STANDARD; PRT; 65 AA.
AC P17159;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Photosystem II reaction center Z protein.
GN PSBZ OR YCF9.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RX MEDLINE=91346714; PubMed=2129400;
RA Erward J.L., Kuntz M., Weil J.H.;
RT "An ORF potentially encoding a 6.5 kDa hydrophobic protein in
RT chloroplasts is also present in the cyanellar genome of Cyanophora
RT paradoxa.";
RL Plant Mol. Biol. 15:779-781(1990).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Jakowitsch W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch W.J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -!- FUNCTION: Controls the interaction of photosystem II (PSII) cores
CC with the light-harvesting antenna (By similarity).
CC -!- SUBCELLULAR LOCATION: Cellular thylakoid membrane; associated with
CC the photosystem II complex (By similarity).
CC -!- SIMILARITY: Belongs to the psbZ family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X51421; CAA35786.1; -
CC EMBL; U30821; AAA81235.1; -
CC PIR; S14712; S14712.
CC HAMAP; MF_00644; -; 1
CC InterPro; IPR002644; Ycf9_struc.
CC Pfam; PF01737; YCF9; 1.
CC Photosynthesis; Photosystem II; Reaction center; Thylakoid;
CC Transmembrane; Cyanelle.
CC TRANSMEM 8 28
CC POTENTIAL.
```

```
FT TRANSMEM 41 61 POTENTIAL.
SQ SEQUENCE 65 AA; 6853 MW; 9FE5CD758B95623C CRC64;

Query Match          2.5%; Score 6; DB 1; Length 65;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPG 176
Db 26 ALASPG 31

RESULT 84
CSRA_XYLPA
ID CSRA_XYLPA STANDARD; PRT; 71 AA.
AC Q9PH21; Q87F42;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carbon storage regulator homolog.
GN CSRA OR XF0125 OR PD0095.
OS Xylella fastidiosa, and
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371, 183190;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=945C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arraya J.E., Baia G.S., Baptista C.S.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bordin S., Bove J.M., Briones M.R.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.B.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=Temeculal / ATCC 700964;
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Moon D.H.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva F.R., Tsai S.M.,
RA Takita M.A., Lemos E.G.M., Machado M.A., de Souza A.A.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Siqueira W.J.,
RA Carver H., Carraro D.M., de Oliveira R.C., Nunes L.R., Kuramae E.E.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.B., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
```

```
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosus";
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- FUNCTION: Could accelerate the degradation of some genes
CC transcripts potentially through selective RNA binding (By
CC similarity).
CC -!- SIMILARITY: Belongs to the csra family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003866; AAF82938.1; ALT_INIT.
DR EMBL; AE012553; AAO27995.1; -.
DR F01; D82844; D82844.
DR HAMAP; MF_00167; -. 1.
DR InterPro; IPR003751; Csra.
DR Pfam; PF02599; Csra; 1.
DR ProDom; PD009007; Csra; 1.
DR TIGRfam; TIGR00402; csra; 1.
KW RNA-binding; Complete proteome.
FT DOMAIN 11 45
FT SEQUENCE 71 AA; 7672 MW; FBA182A40CE72AD CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 RGDGSG 11
DB 64 RGDGSG 69
RESULT 85
C555_METCA STANDARD; PRT; - 96 AA.
AC P04369;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytochrome c-555 (C555).
OS Methylococcus capsulatus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Methylococcales;
OC Methylococcaceae; Methylococcus.
OX NCBI_TaxID=414;
RN [1]
RP SEQUENCE.
RC STRAIN=Bath / NCIB 11132;
RX MEDLINE=86158741; PubMed=3006666;
RA Ambler R.P., Dalton H., Meyer T.E., Bartsch R.G., Kamen M.D.;
RT "The amino acid sequence of cytochrome c-555 from the
RT methane-oxidizing bacterium Methylococcus capsulatus.";
RL Biochem. J. 233:333-337(1986).
DR F01; A23321; CCMPS5.
DR InterPro; IPR003088; Cyt_C1.
DR InterPro; IPR000345; CytC_heme_BS.
DR Pfam; PF00034; Cytochrome C; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
KW Electron transport; Photosynthesis; Heme.
FT BINDING 19 19 HEME (COVALENT).
FT BINDING 22 22 HEME (COVALENT).
FT METAL 23 23 IRON (HEME AXIAL LIGAND).
FT METAL 59 59 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 96 AA; 10506 MW; 17B5DCF79535A585 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 96;
```

```
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 226 AGELEK 231
DB 89 AGELEK 94
|||||
RESULT 86
RSN_MOUSE STANDARD; PRT; 114 AA.
ID RSN_MOUSE
AC Q99P87;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Resistin precursor (Cysteine-rich secreted protein FIZZ3) (Adipose
DE tissue-specific secretory factor) (ADSF) (Adipose-specific cysteine-
DE rich secreted protein A12-alpha).
GN RTN OR FIZZ3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=21069045; PubMed=11201732;
RA Steppan C.M., Bailey S.T., Bhat S., Brown E.J., Banerjee R.R.,
RA Wright C.M., Patel H.R., Ahima R.S., Lazar M.A.;
RT "The hormone resistin links obesity to diabetes.";
RL Nature 409:307-312(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Rajala M.W., Scherer P.E.;
RT "Identification of a novel adipose-specific cysteine-rich secreted
RT protein.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Matovina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SUBUNIT.
RX MEDLINE=21336653; PubMed=11358969;
RA Banerjee R.R., Lazar M.A.;
RT "Dimerization of resistin and resistin-like molecules is determined by
RT a single cysteine.";
RL J. Biol. Chem. 276:25970-25973(2001).
CC -!- FUNCTION: Hormone that seems to suppress insulin ability to
CC stimulate glucose uptake into adipose cells. Potentially
CC links obesity to diabetes.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed in white but not brown adipose
```

tissue in a variety of organs.
-!- PTM: 5 disulfide bonds are present (probable).
-!- SIMILARITY: Belongs to the resistin/FIZZ family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF323080; AAG59823.1; -;
DR EMBL; AF290870; AAK3102.1; -;
DR EMBL; BC051196; AAH51196.1; -;
DR MGD; MGI:1888506; Ratt.
DR GO; GO:0005576; C:extracellular; IDA.
DR Hormone; Signal; Diabetes mellitus; Obesity.
FT SIGNAL 1 20
FT CHAIN 21 114 RESISTIN.
FT DISULFID 26 26 INTERCHAIN (PROBABLE).
SQ SEQUENCE 114 AA; 12491 MW; D44930E51D3F22C8 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SCPEOT 38
Db 66 SOPEOT 71

RESULT 87
ID CN4D MOUSE STANDARD; PRT; 116 AA.
AC Q01063;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CAMP-specific 3',5'-cyclic phosphodiesterase 4D (EC 3.1.4.17) (DPDE3) (fragment).
DE (fragment).
GN PDB4D.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92406782; PubMed=1326532;
RA Repaske D.R., Swinnen J.V., Jin S.L.C., van Wyk J.J., Conti M.;
RT "A polymerase chain reaction strategy to identify and clone cyclic nucleotide phosphodiesterase cDNAs. Molecular cloning of the cDNA encoding the 63-kDa calmodulin-dependent phosphodiesterase.";
RT J. Biol. Chem. 267:18683-18688 (1992).
RL J. Biol. Chem. 267:18683-18688 (1992).
CC -!- CATALYTIC ACTIVITY: Adenosine 3',5'-cyclic phosphate + H(2)O = adenosine 5'-phosphate.
CC -!- ENZYME REGULATION: Inhibited by rolipram.
CC -!- PATHWAY: Cyclic nucleotide metabolism.
CC -!- SIMILARITY: Belongs to the cyclic nucleotide phosphodiesterase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M94541; AAA37368.1; -;
DR MGD; MGI:99555; Pde4d.
DR InterPro; IPR002073; PDEase.
DR Pfam; PF00233; PDEase; 1.

DR PRINTS; PR00387; PD1ESTERASE1.
DR PROSITE; PS00126; PDEASE_I; 1.
KW Hydrolase; CAMP; Alternative splicing; Multigene family.
FT NON_TER 1
FT NON_TER 116 116
SQ SEQUENCE 116 AA; 12928 MW; D217AF0F432611B2 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 LSTPAL 102
Db 18 LSTPAL 23

RESULT 88
ID Y128 SYNTP6 STANDARD; PRT; 119 AA.
AC P05677;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 12.8 kDa protein in 16S rRNA gene region.
OS Synecococcus sp. (strain POC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
CX NCBI_TaxID=1139;
RN [1]
RP SEQUENCE FROM N.A.
RA Kumano M., Tomioka N., Shinozaki K., Sugitara M.;
RT "Analysis of the promoter region in the rRNA operon from a blue-green alga, Anacystis nidulans 6301.";
RL Mol. Gen. Genet. 202:173-178 (1986).
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03538; CAA27241.1; -;
DR PIR; S10914; S10914.
DR InterPro; IPR004843; M-pdestrase.
DR Pfam; PF00149; Metallophos; 1.
KW Hypothetical protein.
SQ SEQUENCE 119 AA; 12867 MW; F1032C437D5DFEF6 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIA 133
Db 73 PAIAIA 78

RESULT 89
ID YMJ0 YEAST STANDARD; PRT; 128 AA.
AC Q04501;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Very hypothetical 15.0 kDa protein in RPM2-TUB1 intergenic region.
OS YML090W.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.


```
RC STRAIN=6288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowmar S., Churhan C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Race P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93 (1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z46660; CAA86648.1; --
DR PIR; S49537; S49637.
DR GenOnline; 142621; --
DR SGD; S0004555; YML090W.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT SEQUENCE 128 AA; 15034 MW; A4AE42CDFA441B6B CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 PSFLFV 50
DB 4 PSFLFV 9
RESULT 90
ID_2_MOUSE STANDARD; PRT; 130 AA.
AC Q99MP3;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calcitonin gene-related peptide II precursor (CGRP-II) (Beta-type
DE CGRP).
DE CALCB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21604266; PubMed=11761712;
RA Thomas P.M., Nasonkin I., Zhang H., Gagel R.F., Cote G.J.;
RT "Structure of the mouse calcitonin/calcitonin gene-related peptide
RT alpha and beta genes."
RL DNA Seq. 12:131-135 (2001).
CC -1- FUNCTION: CGRP induces vasodilatation. It dilates a variety of
CC vessels including the coronary, cerebral and systemic vasculature.
CC Its abundance in the CNS also points toward a neurotransmitter or
CC neuromodulator role (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the calcitonin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF325526; AAK16431.1; --
DR EMBL; AF325524; AAK16431.1; JOINED.
DR MGD; MGI:2151254; Calcb.
DR InterPro; IPR001693; Calcitonin-like.
DR InterPro; IPR002163; Calcitonin B.
DR Pfam; PF00214; Calc. CGRP IAPP; 1.
DR PRINTS; PR00817; CALCITONINB.
DR SMART; SMO0133; CALCITONIN; 1.
DR PROSITE; PS00258; CALCITONIN; 1.
KW Cleavage on pair of basic residues; Amidation; Hormone; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 82 BY SIMILARITY.
FT PEPTIDE 84 120 CALCITONIN GENE-RELATED PEPTIDE II.
FT PROPEP 127 130 BY SIMILARITY.
FT DISULFID 85 90 BY SIMILARITY.
FT MOD_RES 120 120 AMIDATION (G-121 PROVIDE AMIDE GROUP) (BY
FT SIMILARITY).
SQ SEQUENCE 130 AA; 14623 MW; 97299244E8F6C536 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 DLGTLG 65
DB 38 DLGTLG 43
RESULT 91
YAEJ_PSEPU STANDARD; PRT; 137 AA.
ID YAEJ_PSEPU
AC P45388;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical 15.2 kDa protein in pcau 3 region.
DE Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PRS2000;
RX MEDLINE=92325057; PubMed=1624453;
RA Parales R.E., Harwood C.S.;
RT "Characterization of the genes encoding beta-ketoadipate: succinyl-
RT coenzyme A transferase in Pseudomonas putida."
RL J. Bacteriol. 174:4657-4666 (1992).
CC -1- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
CC factor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88763; -- NOT ANNOTATED CDS.
DR InterPro; IPR000352; Pep_rel_factor_I.
DR Pfam; PF00472; RF-1; 1.
DR PROSITE; PS00745; RF_PROK_I; 1.
KW Hypothetical protein.
SQ SEQUENCE 137 AA; 15181 MW; BDAF2B68986A2EC CRC64;
Query Match 2.5%; Score 6; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 222 STVAG 227
```

CC -1- FUNCTION: Troponin is the central regulatory protein of striated muscle.


```
RT of stone formation.";
RL EMBO J. 15:2678-2684(1996).
RN [15]
RP X-RAY CRYSTALLOGRAPHY (1.30 ANGSTROMS) OF 23-166.
RX MEDLINE=20092874; PubMed=10625646;
RA Gerbaud V., Pignol D., Loret E., Bertrand J.A., Berland Y.,
RA Fontecilla-Camps J.C., Canselier J.P., Gabas N., Verdier J.M.;
RT Mechanism of calcite crystal growth inhibition by the N-terminal
RT undecapeptide of lithostathine.";
RL J. Biol. Chem. 275:1057-1064(2000).
RN [16]
RN STRUCTURE BY NMR OF 34-164.
RX MEDLINE=97120677; PubMed=8961348;
RA Patard L., Stoven V., Gharib B., Bontems F., Lallemand J.-Y.,
RA de Reggi M.;
RT "What function for human lithostathine?: structural investigations by
RT three-dimensional structure modeling and high-resolution NMR
RT spectroscopy.";
RL Protein Eng. 9:949-957(1996).
CC -!- FUNCTION: Might act as an inhibitor of spontaneous calcium
CC carbonate precipitation. May be associated with neuronal
CC sprouting in brain, and with brain and pancreas regeneration.
CC -!- TISSUE SPECIFICITY: In pancreatic acinar cells and, in lower
CC levels, in brain.
CC -!- DEVELOPMENTAL STAGE: High expression levels in fetal and infant
CC brains; much lower in adult brains.
CC -!- DISEASE: Alzheimer's disease and Down's syndrome patients show
CC enhanced expression of PSP-related transcripts and intraneuronal
CC accumulation of PSP-like proteins in their brains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27130; AAA60546.1; -
DR EMBL; M27189; AAA60545.1; -
DR EMBL; M18963; AAA36558.1; -
DR EMBL; J05412; AAA36559.1; -
DR EMBL; AF172331; AAD51330.1; -
DR EMBL; BC005350; AAA05350.1; -
DR PIR; A35197; RGHU1A.
DR PIR; A45751; A45751.
DR PDB; 1LIT; 11-JAN-97.
DR PDB; 1QDD; 24-JAN-01.
DR Genew; HGNC:19951; REG1A.
DR MIM; 167770; -
DR MIM; 167800; -
DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
DR InterPro; IPR002353; Antifreeze1.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancreatis_ac.
DR Pfam; PF00059; lectin_C; 1.
DR PRINTS; PR00356; ANTIFREEZE1.
DR PRINTS; PR01504; PNCRATISAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
KW Glycoprotein; Signal; Alzheimer's disease; Down's syndrome; Lectin;
KW 3D-structure; Pyrrolidone carboxylic acid.
FT SIGNAL 1 22
Query Match 2.5%; Score 6; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 SCEPT 38
DB 35 SCEPT 40
```

```
RESULT 96
LITE_HUMAN
ID LITE_HUMAN STANDARD; PRT; 166 AA.
AC P48304;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lithostathine 1 beta precursor (Regenerating protein I beta).
GN REG1B OR REG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RS TISSUE=Pancreas;
RX MEDLINE=94153997; PubMed=8110835;
RA Morizumi S., Watanabe T., Unno M., Nakagawara K.I., Suzuki Y.,
RA Miyashita H., Yonekura H., Okamoto H.;
RT "Isolation, structural determination and expression of a novel reg
RT gene, human regi beta.";
RL Biochim. Biophys. Acta 1217:199-202(1994).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=93351647; PubMed=8348956;
RA Bartoli C., Gharib B., Giorgi D., Sansonetti A., Dagorn J.-C.,
RA Berge-Lefranc U.;
RT "A gene homologous to the reg gene is expressed in the human
RT pancreas.";
RL FEBS Lett. 327:289-293(1993).
RN [3]
RN SEQUENCE FROM N.A.
RS TISSUE=Pancreas;
RX MEDLINE=2238257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Pearce C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RN CARBOHYDRATE-LINKAGE SITE.
RX MEDLINE=95331286; PubMed=7607222;
RA De Reggi M., Capon C., Gharib B., Wieruszski J.M., Michel R.,
RA Fournet B.;
RT "The glycan moiety of human pancreatic lithostathine. Structure
RT characterization and possible pathophysiological implications.";
RL Eur. J. Biochem. 230:503-510(1995).
CC -!- FUNCTION: Might act as an inhibitor of spontaneous calcium
CC carbonate precipitation. May be associated with neuronal sprouting
CC in brain, and with brain and pancreas regeneration.
CC -!- PTM: ALL O-LINKED GLYCANS CONSIST OF GAL-GLCNAC-GAL-GALNAC
CC TETRASACCHARIDE CORE AND GET ELONGATED (MICROHETEROGENEITY).
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D17291; BAA04124.1; -
CC EMBL; D16816; BAA04091.1; -
CC EMBL; L08010; AAA18204.1; -
CC EMBL; BC027895; AAH27895.1; -
CC PIR; S34591; RGHU1B.
CC HSSP; P05451; LLIT.
CC Genew; HGNC:9952; REGIB.
CC MIM; 167771; -
CC GO; GO:0008283; P:cell proliferation; TAS.
CC InterPro; IPR001304; LECTIN_C.
CC InterPro; IPR003990; Pancreatins_ac.
CC Pfam; PF00059; lectin_c; 1.
CC PRINTS; PR01504; PNCREATITSAP.
CC SMART; SM00334; CLECT; 1.
CC PROSITE; PS00615; C-TYPE LECTIN 1; 1.
CC PROSITE; PS0041; C-TYPE LECTIN 2; 1.
CC Glycoprotein; Signal; LECTIN; Pyroglutamate carboxylic acid.
CC SIGNAL 1 22 BY SIMILARITY.
CC CHAIN 23 166 LITHOSTATHINE 1 BETA.
CC DOMAIN 34 164 C-TYPE LECTIN (LONG FORM).
CC MOD_RES 23 23 PYROGLUTAMATE CARBOXYLIC ACID
CC (BY SIMILARITY).
CC CARBOHYD 27 27 O-LINKED (GALNAc...) (MUCIN TYPE).
CC DISULFID 36 47 BY SIMILARITY.
CC DISULFID 64 162 BY SIMILARITY.
CC DISULFID 137 154 BY SIMILARITY.
CC SEQUENCE 166 AA; 18665 MW; D1DC20E11AE5DDE8 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SCEPT 38
DB 35 SCEPT 40

RESULT 97
VB03 VACCV STANDARD; PRT; 167 AA.
AC Q0126;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Protein B3.
OS B3.
GN Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
FN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91259063; PubMed=2045793;
RA Smith G.L., Chan Y.S., Howard S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT the right inverted terminal repeat";
RL J. Gen. Virol. 72:1349-1376(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D11079; BAA01833.1; -

DR PIR; JQ1797; JQ1797.
SQ SEQUENCE 167 AA; 19410 MW; 82AF46891A7768D7 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 167;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 IPSTVK 225
DB 39 IPSTVK 44

RESULT 98
RECU UREPA STANDARD; PRT; 170 AA.
ID RECU UREPA
AC Q9PQJ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Recombination protein U homolog.
GN RSCU OR U297.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RC MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum";
RL Nature 407:757-762(2000).
CC -!- FUNCTION: Required for DNA repair and intramolecular
CC recombination. Seems also to be required for chromosome
CC segregation (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the recU family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE002127; AAF30706.1; -
CC HAMAP; MF 00130; -; 1.
DR InterPro; IPR004612; RecU.
DR Pfam; PF03838; RecU; 1.
DR DNA repair; DNA recombination; Complete proteome.
SQ SEQUENCE 170 AA; 20527 MW; 892315A8C36933DD CRC64;

Query Match 2.5%; Score 6; DB 1; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 SGFSFL 48
DB 99 SGFSFL 104

RESULT 99
RL10 THEAC STANDARD; PRT; 176 AA.
ID RL10 THEAC
AC Q9HJB3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L10e.
GN RPL10E OR TAI057.
OS Thermoplasma acidophilum.

```
OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatatales;
OC Thermoplasmatataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Grail W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Newes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
CC -1- SIMILARITY: Belongs to the L10e family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR HAMAP; MF 00448; -; 1.
DR InterPro; IPR001197; Ribosomal L10E.
DR Pfam; PF00826; Ribosomal L10e; 1.
DR PROSITE; PS01257; RIBOSOMAL_L10E; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 176 AA; 19377 MW; E284F9F14A187B46 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 FVQGNQ 54
DB 34 FVQGNQ 39
|||||

RESULT 101
INAA BOVIN STANDARD; PRT; 189 AA.
AC P05007;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Interferon alpha-A precursor.
GN IFNAA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85182698; PubMed=3886658;
RA Velan B., Cohen S., Grosfeld H., Leitner M., Shafferman A.;
RT "Bovine interferon alpha genes. Structure and expression."
RL J. Biol. Chem. 260:5458-5504(1985).
CC -1- FUNCTION: PRODUCED BY MACROPHAGES, IFN-ALPHA HAVE ANTIVIRAL
CC ACTIVITIES. INTERFERON STIMULATES THE PRODUCTION OF TWO ENZYMES:
CC A PROTEIN KINASE AND AN OLIGODENYLATE SYNTHETASE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the alpha/beta interferon family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Db          95 GSPATW 100

RESULT 102
SODN STAAM
ID SODN STAAM STANDARD; PRT; 199 AA.
AC Q99X82;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Superoxide dismutase [Mn/Fe] 2 (EC 1.15.1.1).
GN SOD2 OR SAV0133 OR SA0128 OR MW0107.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcaceae;
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RA "whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=WM2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 manganese or iron ion per subunit (By
CC similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003358; BAB56295.1; -
CC EMBL; AP003129; BAB41348.1; -
CC EMBL; AP004822; BAB93972.1; -
CC PIR; A89774; A89774.
CC HSP; P09214; 1MNG.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC PRODOM; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Iron; Complete proteome.
KW OXIDOREDUCTASE; SOD_MN; 1.
FT METAL 27 27 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 81 81 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 161 161 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 161 161 MANGANESE OR IRON (BY SIMILARITY).

FT METAL 165 165 MANGANESE OR IRON (BY SIMILARITY).
SQ SEQUENCE 199 AA; 23041 MW; 388566FB9943C635 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 199;
Best Local Similarity 100.0%; Pred.No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYI 119
Db 14 ALEPYI 19

RESULT 103
SODM STRPN
ID SODM STRPN STANDARD; PRT; 200 AA.
AC Q59949; Q33757; Q54268; Q54269; Q9R3B8; Q9R3B8; Q9S176;
AC Q9S177; Q9S447;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn] (EC 1.15.1.1).
GN SODA OR SP0766 OR SP0674.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=1313, 171101;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=D39 / NCTC 7466 / Serotype 2;
RX MEDLINE=20231823; PubMed=10768978;
RA Yesilkaya H., Kadioglu A., Gingles N., Alexander J.E., Mitchell T.J.,
RA Andrew P.W.;
RT "Role of manganese-containing superoxide dismutase in oxidative stress
RT and virulence of Streptococcus pneumoniae.";
RL Infect. Immun. 68:2819-2826(2000).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., White O., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA McKay L.A., Whittam M., Salzbarg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21423245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA Dehoff B.S., Essem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.J., Legace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nickas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
RN [4]
SEQUENCE OF 17-161 FROM N.A.
RP STRAIN=NEW667;
RX MEDLINE=96045282; PubMed=7557308;
RA Poyart C., Berche P., Trieu-Cuot P.;
RT "Characterization of superoxide dismutase genes from Gram-positive
RT bacteria by polymerase chain reaction using degenerate primers.";
```

RL FEMS Microbiol. Lett. 131:41-45(1995).

RN [5]

RP SEQUENCE OF 17-161 FROM N.A.

RC STRAIN=Various strains;

RX MEDLINE=98092214; PubMed=9431917;

RY Poyart C., Quesne G., Coulon S., Berche P., Trieu-Cuot P.;

RT "Identification of streptococci to species level by sequencing the

RT gene encoding the manganese-dependent superoxide dismutase.";

RL J. Clin. Microbiol. 36:41-47(1998).

RN [6]

RP SEQUENCE OF 32-153 FROM N.A.

RC STRAIN=653, 661, 872, 1293, 1454, 1510, 1565, 1639, 3051, 3203,

RC GTC361T / NCTC 7465T, YK-5, YK-11, YK-12, YK-14, and YK-20;

RX MEDLINE=99445202; PubMed=10517614;

RY Kawamura Y., Whitley R.A., Shu S.E., Ezaki T., Hardie J.M.;

RT "Genetic approaches to the identification of the mitis group within

RT the genus Streptococcus.";

RL Microbiology 145:260S-2613(1999).

CC -!- FUNCTION: Destroys radicals which are normally produced within the

CC cells and which are toxic to biological systems. May play a

CC critical role against oxidative stress, affecting both the survival

CC and the virulence of *S. pneumoniae*.

CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).

CC -!- SUBUNIT: Homodimer (By similarity).

CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase

CC family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF162664; AA050778.1; -

DR EMBL; AB007384; AA074904.1; -

DR EMBL; AB008445; AA099478.1; -

DR EMBL; Z49246; CA089213.1; -

DR EMBL; Z95914; CA089367.1; -

DR EMBL; Z99200; CA016344.1; -

DR EMBL; Z99201; CA016345.1; -

DR EMBL; Z99202; CA016346.1; -

DR EMBL; Z99203; CA016347.1; -

DR EMBL; Z99204; CA016348.1; -

DR EMBL; Z99205; CA016349.1; -

DR EMBL; Z99206; CA016350.1; -

DR EMBL; AB021544; BA085492.1; -

DR EMBL; AB021545; BA085493.1; -

DR EMBL; AB021546; BA085494.1; -

DR EMBL; AB021547; BA085495.1; -

DR EMBL; AB021548; BA085496.1; -

DR EMBL; AB021549; BA085497.1; -

DR EMBL; AB021550; BA085498.1; -

DR EMBL; AB021551; BA085499.1; -

DR EMBL; AB021552; BA085500.1; -

DR EMBL; AB021553; BA085501.1; -

DR EMBL; AB021554; BA085502.1; -

DR EMBL; AB021555; BA085503.1; -

DR EMBL; AB021556; BA085504.1; -

DR EMBL; AB021557; BA085505.1; -

DR EMBL; AB021558; BA085506.1; -

DR EMBL; AB021559; BA085507.1; -

DR EMBL; AB021560; BA085508.1; -

DR EMBL; AB021561; BA085509.1; -

DR EMBL; AB021562; BA085510.1; -

DR EMBL; AB021563; BA085511.1; -

DR EMBL; AB021564; BA085512.1; -

DR EMBL; AB021565; BA085513.1; -

DR EMBL; AB021566; BA085514.1; -

DR EMBL; AB021567; BA085515.1; -

DR EMBL; AB021568; BA085516.1; -

DR EMBL; AB021569; BA085517.1; -

DR EMBL; AB021570; BA085518.1; -

DR EMBL; AB021571; BA085519.1; -

DR EMBL; AB021572; BA085520.1; -

DR EMBL; AB021573; BA085521.1; -

DR EMBL; AB021574; BA085522.1; -

DR EMBL; AB021575; BA085523.1; -

DR EMBL; AB021576; BA085524.1; -

DR EMBL; AB021577; BA085525.1; -

DR EMBL; AB021578; BA085526.1; -

DR EMBL; AB021579; BA085527.1; -

DR EMBL; AB021580; BA085528.1; -

DR EMBL; AB021581; BA085529.1; -

DR EMBL; AB021582; BA085530.1; -

DR EMBL; AB021583; BA085531.1; -

DR EMBL; AB021584; BA085532.1; -

DR EMBL; AB021585; BA085533.1; -

DR EMBL; AB021586; BA085534.1; -

DR EMBL; AB021587; BA085535.1; -

DR EMBL; AB021588; BA085536.1; -

DR EMBL; AB021589; BA085537.1; -

DR EMBL; AB021590; BA085538.1; -

DR EMBL; AB021591; BA085539.1; -

DR EMBL; AB021592; BA085540.1; -

DR EMBL; AB021593; BA085541.1; -

DR EMBL; AB021594; BA085542.1; -

DR EMBL; AB021595; BA085543.1; -

DR EMBL; AB021596; BA085544.1; -

DR EMBL; AB021597; BA085545.1; -

DR EMBL; AB021598; BA085546.1; -

DR EMBL; AB021599; BA085547.1; -

DR EMBL; AB021600; BA085548.1; -

DR EMBL; AB021601; BA085549.1; -

DR EMBL; AB021602; BA085550.1; -

DR EMBL; AB021603; BA085551.1; -

DR EMBL; AB021604; BA085552.1; -

DR EMBL; AB021605; BA085553.1; -

DR EMBL; AB021606; BA085554.1; -

DR EMBL; AB021607; BA085555.1; -

DR EMBL; AB021608; BA085556.1; -

DR EMBL; AB021609; BA085557.1; -

DR EMBL; AB021610; BA085558.1; -

DR EMBL; AB021611; BA085559.1; -

DR EMBL; AB021612; BA085560.1; -

DR EMBL; AB021613; BA085561.1; -

DR EMBL; AB021614; BA085562.1; -

DR EMBL; AB021615; BA085563.1; -

DR EMBL; AB021616; BA085564.1; -

DR EMBL; AB021617; BA085565.1; -

DR EMBL; AB021618; BA085566.1; -

DR EMBL; AB021619; BA085567.1; -

DR EMBL; AB021620; BA085568.1; -

DR EMBL; AB021621; BA085569.1; -

DR EMBL; AB021622; BA085570.1; -

DR EMBL; AB021623; BA085571.1; -

DR EMBL; AB021624; BA085572.1; -

DR EMBL; AB021625; BA085573.1; -

DR EMBL; AB021626; BA085574.1; -

DR EMBL; AB021627; BA085575.1; -

DR EMBL; AB021628; BA085576.1; -

DR EMBL; AB021629; BA085577.1; -

DR EMBL; AB021630; BA085578.1; -

DR EMBL; AB021631; BA085579.1; -

DR EMBL; AB021632; BA085580.1; -

DR EMBL; AB021633; BA085581.1; -

DR EMBL; AB021634; BA085582.1; -

DR EMBL; AB021635; BA085583.1; -

DR EMBL; AB021636; BA085584.1; -

DR EMBL; AB021637; BA085585.1; -

DR EMBL; AB021638; BA085586.1; -

DR EMBL; AB021639; BA085587.1; -

DR EMBL; AB021640; BA085588.1; -

DR EMBL; AB021641; BA085589.1; -

DR EMBL; AB021642; BA085590.1; -

DR EMBL; AB021643; BA085591.1; -

DR EMBL; AB021644; BA085592.1; -

DR EMBL; AB021645; BA085593.1; -

DR EMBL; AB021646; BA085594.1; -

DR EMBL; AB021647; BA085595.1; -

DR EMBL; AB021648; BA085596.1; -

DR EMBL; AB021649; BA085597.1; -

DR EMBL; AB021650; BA085598.1; -

DR EMBL; AB021651; BA085599.1; -

DR EMBL; AB021652; BA085600.1; -

DR EMBL; AB021653; BA085601.1; -

DR EMBL; AB021654; BA085602.1; -

DR EMBL; AB021655; BA085603.1; -

DR EMBL; AB021656; BA085604.1; -

DR EMBL; AB021657; BA085605.1; -

DR EMBL; AB021658; BA085606.1; -

DR EMBL; AB021659; BA085607.1; -

DR EMBL; AB021660; BA085608.1; -

DR EMBL; AB021661; BA085609.1; -

DR EMBL; AB021662; BA085610.1; -

DR EMBL; AB021663; BA085611.1; -

DR EMBL; AB021664; BA085612.1; -

DR EMBL; AB021665; BA085613.1; -

DR EMBL; AB021666; BA085614.1; -

DR EMBL; AB021667; BA085615.1; -

DR EMBL; AB021668; BA085616.1; -

DR EMBL; AB021669; BA085617.1; -

DR EMBL; AB021670; BA085618.1; -

DR EMBL; AB021671; BA085619.1; -

DR EMBL; AB021672; BA085620.1; -

DR EMBL; AB021673; BA085621.1; -

DR EMBL; AB021674; BA085622.1; -

DR EMBL; AB021675; BA085623.1; -

DR EMBL; AB021676; BA085624.1; -

DR EMBL; AB021677; BA085625.1; -

DR EMBL; AB021678; BA085626.1; -

DR EMBL; AB021679; BA085627.1; -

DR EMBL; AB021680; BA085628.1; -

DR EMBL; AB021681; BA085629.1; -

DR EMBL; AB021682; BA085630.1; -

DR EMBL; AB021683; BA085631.1; -

DR EMBL; AB021684; BA085632.1; -

DR EMBL; AB021685; BA085633.1; -

DR EMBL; AB021686; BA085634.1; -

DR EMBL; AB021687; BA085635.1; -

DR EMBL; AB021688; BA085636.1; -

DR EMBL; AB021689; BA085637.1; -

DR EMBL; AB021690; BA085638.1; -

DR EMBL; AB021691; BA085639.1; -

DR EMBL; AB021692; BA085640.1; -

DR EMBL; AB021693; BA085641.1; -

DR EMBL; AB021694; BA085642.1; -

DR EMBL; AB021695; BA085643.1; -

DR EMBL; AB021696; BA085644.1; -

DR EMBL; AB021697; BA085645.1; -

DR EMBL; AB021698; BA085646.1; -

DR EMBL; AB021699; BA085647.1; -

DR EMBL; AB021700; BA085648.1; -

DR EMBL; AB021701; BA085649.1; -

DR EMBL; AB021702; BA085650.1; -

DR EMBL; AB021703; BA085651.1; -

DR EMBL; AB021704; BA085652.1; -

DR EMBL; AB021705; BA085653.1; -

DR EMBL; AB021706; BA085654.1; -

DR EMBL; AB021707; BA085655.1; -

DR EMBL; AB021708; BA085656.1; -

DR EMBL; AB021709; BA085657.1; -

DR EMBL; AB021710; BA085658.1; -

DR EMBL; AB021711; BA085659.1; -

DR EMBL; AB021712; BA085660.1; -

DR EMBL; AB021713; BA085661.1; -

DR EMBL; AB021714; BA085662.1; -

DR EMBL; AB021715; BA085663.1; -

DR EMBL; AB021716; BA085664.1; -

DR EMBL; AB021717; BA085665.1; -

DR EMBL; AB021718; BA085666.1; -

DR EMBL; AB021719; BA085667.1; -

DR EMBL; AB021720; BA085668.1; -

DR EMBL; AB021721; BA085669.1; -

DR EMBL; AB021722; BA085670.1; -

DR EMBL; AB021723; BA085671.1; -

DR EMBL; AB021724; BA085672.1; -

DR EMBL; AB021725; BA085673.1; -

DR EMBL; AB021726; BA085674.1; -

DR EMBL; AB021727; BA085675.1; -

DR EMBL; AB021728; BA085676.1; -

DR EMBL; AB021729; BA085677.1; -

DR EMBL; AB021730; BA085678.1; -

DR EMBL; AB021731; BA085679.1; -

DR EMBL; AB021732; BA085680.1; -

DR EMBL; AB021733; BA085681.1; -

DR EMBL; AB021734; BA085682.1; -

DR EMBL; AB021735; BA085683.1; -

DR EMBL; AB021736; BA085684.1; -

DR EMBL; AB021737; BA085685.1; -

DR EMBL; AB021738; BA085686.1; -

DR EMBL; AB021739; BA085687.1; -

DR EMBL; AB021740; BA085688.1; -

DR EMBL; AB021741; BA085689.1; -

DR EMBL; AB021742; BA085690.1; -

DR EMBL; AB021743; BA085691.1; -

DR EMBL; AB021744; BA085692.1; -

DR EMBL; AB021745; BA085693.1; -

DR EMBL; AB021746; BA085694.1; -

DR EMBL; AB021747; BA085695.1; -

DR EMBL; AB021748; BA085696.1; -

DR EMBL; AB021749; BA085697.1; -

DR EMBL; AB021750; BA085698.1; -

DR EMBL; AB021751; BA085699.1; -

DR EMBL; AB021752; BA085700.1; -

DR EMBL; AB021753; BA085701.1; -

DR EMBL; AB021754; BA085702.1; -

DR EMBL; AB021755; BA085703.1; -

DR EMBL; AB021756; BA085704.1; -

DR EMBL; AB021757; BA085705.1; -

DR EMBL; AB021758; BA085706.1; -

DR EMBL; AB021759; BA085707.1; -

DR EMBL; AB021760; BA085708.1; -

DR EMBL; AB021761; BA085709.1; -

DR EMBL; AB021762; BA085710.1; -

DR EMBL; AB021763; BA085711.1; -

DR EMBL; AB021764; BA085712.1; -

DR EMBL; AB021765; BA085713.1; -

DR EMBL; AB021766; BA085714.1; -

DR EMBL; AB021767; BA085715.1; -

DR EMBL; AB021768; BA085716.1; -

DR EMBL; AB021769; BA085717.1; -

DR EMBL; AB021770; BA085718.1; -

DR EMBL; AB021771; BA085719.1; -

DR EMBL; AB021772; BA085720.1; -

DR EMBL; AB021773; BA085721.1; -

DR EMBL; AB021774; BA085722.1; -

DR EMBL; AB021775; BA085723.1; -

DR EMBL; AB021776; BA085724.1; -

DR EMBL; AB021777; BA085725.1; -

DR EMBL; AB021778; BA085726.1; -

DR EMBL; AB021779; BA085727.1; -

DR EMBL; AB021780; BA085728.1; -

DR EMBL; AB021781; BA085729.1; -

DR EMBL; AB021782; BA085730.1; -

DR EMBL; AB021783; BA085731.1; -

DR EMBL; AB021784; BA085732.1; -

DR EMBL; AB021785; BA085733.1; -

DR EMBL; AB021786; BA085734.1; -

DR EMBL; AB021787; BA085735.1; -

DR EMBL; AB021788; BA085736.1; -

DR EMBL; AB021789; BA085737.1; -

DR EMBL; AB021790; BA085738.1; -

DR EMBL; AB021791; BA085739.1; -

DR EMBL; AB021792; BA085740.1; -

DR EMBL; AB021793; BA085741.1; -

DR EMBL; AB021794; BA085742.1; -

DR EMBL; AB021795; BA085743.1; -

DR EMBL; AB021796; BA085744.1; -

DR EMBL; AB021797; BA085745.1; -

DR EMBL; AB021798; BA085746.1; -

DR EMBL; AB021799; BA085747.1; -

DR EMBL; AB021800; BA085748.1; -

DR EMBL; AB021801; BA085749.1; -

DR EMBL; AB021802; BA085750.1; -

DR EMBL; AB021803; BA085751.1; -

DR EMBL; AB021804; BA085752.1; -

DR EMBL; AB021805; BA085753.1; -

DR EMBL; AB021806; BA085754.1; -

DR EMBL; AB021807; BA085755.1; -

DR EMBL; AB021808; BA085756.1; -

DR EMBL; AB021809; BA085757.1; -

DR EMBL; AB021810; BA085758.1; -

DR EMBL; AB021811; BA085759.1; -

DR EMBL; AB021812; BA085760.1; -

DR EMBL; AB021813; BA085761.1; -

DR EMBL; AB021814; BA085762.1; -

DR EMBL; AB021815; BA085763.1; -

DR EMBL; AB021816; BA085764.1; -

DR EMBL; AB021817; BA085765.1; -

DR EMBL; AB021818; BA085766.1; -

DR EMBL; AB021819; BA085767.1; -

DR EMBL; AB021820; BA085768.1; -

DR EMBL; AB021821; BA085769.1; -

DR EMBL; AB021822; BA085770.1; -

DR EMBL; AB021823; BA085771.1; -

DR EMBL; AB021824; BA085772.1; -

DR EMBL; AB021825; BA085773.1; -

DR EMBL; AB021826; BA085774.1; -

DR EMBL; AB021827; BA085775.1; -

DR EMBL; AB021828; BA085776.1; -

DR EMBL; AB021829; BA085777.1; -

DR EMBL; AB021830; BA085778.1; -

DR EMBL; AB021831; BA085779.1; -

DR EMBL; AB021832; BA085780.1; -

DR EMBL; AB021833; BA085781.1; -

DR EMBL; AB021834; BA085782.1; -

DR EMBL; AB021835; BA085783.1; -

DR EMBL; AB021836; BA085784.1; -

DR EMBL; AB021837; BA085785.1; -

DR EMBL; AB021838; BA085786.1; -

DR EMBL; AB021839; BA085787.1; -

DR EMBL; AB021840; BA085788.1; -

DR EMBL; AB021841; BA085789.1; -

DR EMBL; AB021842; BA085790.1; -

DR EMBL; AB021843; BA085791.1; -

DR EMBL; AB021844; BA085792.1; -

DR EMBL; AB021845; BA085793.1; -

DR EMBL; AB021846; BA085794.1; -

DR EMBL; AB021847; BA085795.1; -

DR EMBL; AB021848; BA085796.1; -

DR EMBL; AB021849; BA085797.1; -

DR EMBL; AB021850; BA085798.1; -

DR EMBL; AB021851; BA085799.1; -

DR EMBL; AB021852; BA085800.1; -

DR EMBL; AB021853; BA085801.1; -

DR EMBL; AB021854; BA085802.1; -

DR EMBL; AB021855; BA085803.1; -

DR EMBL; AB021856; BA085804.1; -

DR EMBL; AB021857; BA085805.1; -

DR EMBL; AB021858; BA085806.1; -

DR EMBL; AB021859; BA085807.1; -

DR EMBL; AB021860; BA085808.1; -

DR EMBL; AB021861; BA085809.1; -

DR EMBL; AB021862; BA085810.1; -

DR EMBL; AB021863; BA085811.1; -

DR EMBL; AB021864; BA085812.1; -

DR EMBL; AB021865; BA085813.1; -

DR EMBL; AB021866; BA085814.1; -

DR EMBL; AB021867; BA085815.1; -

DR EMBL; AB021868; BA085816.1; -

DR EMBL; AB021869; BA085817.1; -

DR EMBL; AB021870; BA085818.1; -

DR EMBL; AB021871; BA085819.1; -

DR EMBL; AB021872; BA085820.1; -

DR EMBL; AB021873; BA085821.1; -

DR EMBL; AB021874; BA085822.1; -

DR EMBL; AB021875; BA085823.1; -

DR EMBL; AB021876; BA085824.1; -

DR EMBL; AB021877; BA085825.1; -

DR EMBL; AB021878; BA085826.1; -

DR EMBL; AB021879; BA085827.1; -

DR EMBL; AB021880; BA085828.1; -

DR EMBL; AB021881; BA085829.1; -

DR EMBL; AB021882; BA085830.1; -

DR EMBL; AB021883; BA085831.1; -

DR EMBL; AB021884; BA085832.1; -

DR EMBL; AB021885; BA085833.1; -

DR EMBL; AB021886; BA085834.1; -

DR EMBL; AB021887; BA085835.1; -

DR EMBL; AB021888; BA085836.1; -

DR EMBL; AB021889; BA085837.1; -

DR EMBL; AB021890; BA085838.1; -

DR EMBL; AB021891; BA085839.1; -

DR EMBL; AB021892; BA085840.1; -

DR EMBL; AB021893; BA085841.1; -

DR EMBL; AB021894; BA085842.1; -

DR EMBL; AB021895; BA085843.1; -

DR EMBL; AB021896; BA085844.1; -

DR EMBL; AB021897; BA085845.1; -

DR EMBL; AB021898; BA085846.1; -

DR EMBL; AB021899; BA085847.1; -

DR EMBL; AB021900; BA085848.1; -

DR EMBL; AB021901; BA085849.1; -

DR EMBL; AB021902; BA085850.1; -

DR EMBL; AB021903; BA085851.1; -

DR EMBL; AB021904; BA085852.1; -

DR EMBL; AB021905; BA085853.1; -

DR EMBL; AB021906; BA085854.1; -

DR EMBL; AB021907; BA085855.1; -

DR EMBL; AB021908; BA085856.1; -

DR EMBL; AB021909; BA085857.1; -

DR EMBL; AB021910; BA085858.1; -

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AL118506; CAC28315.1; --
DR Genew; HGNC:16129; C200xf136.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PSS0105; SAM_DOMAIN; 1.
KW Hypothetical protein.
FT DOMAIN 118 184 SAM.
SQ SEQUENCE 202 AA; 22770 MW; DF2678F090A3B946 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 ITGRAL 115
|||
DB 147 ITGRAL 152

RESULT 105
SODM_STRMU STANDARD; PRT; 202 AA.
ID SODM_STRMU STANDARD; PRT; 202 AA.
AC P09738; Q59791;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).
GN SODA OR SOD OR SMO.629.

OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=92332426; PubMed=1321118;
RA Nakayama K.;
RT "Nucleotide sequence of Streptococcus mutans superoxide dismutase
RT gene and isolation of insertion mutants.";
RL J. Bacteriol. 174:4928-4934(1992).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

RN [3]
RP SEQUENCE OF 1-22.
RX MEDLINE=56250886; PubMed=3722201;
RA Martin M.E., Byers B.R., Olson M.O.J., Salin M.L., Arceneaux J.E.L.,
RA Tolbert C.;
RT "A Streptococcus mutans superoxide dismutase that is active with
RT either manganese or iron as a cofactor.";
RL J. Biol. Chem. 261:9361-9367(1986).

CC -/- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.

CC -/- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -/- COFACTOR: Binds 1 manganese or iron ion per subunit (By
CC similarity).

CC -/- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D01037; BAB86970.1; --
DR EMBL; AB014906; AAN58363.1; --
DR PIR; A42710; A42710.
DR HSP; P00448; IVEW.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe.1.
DR Pfam; PF02777; sodfe.C.1.
DR PRINTS; PR01703; MNSODISMTASE.
DR PRODOM; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.

KW Oxidoreductase; Metal-binding; Manganese; Iron; Complete proteome.
FT INIT MET 0
FT METAL 26 26 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 80 80 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 162 162 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 166 166 MANGANESE OR IRON (BY SIMILARITY).
FT CONFLICT 3 3 L -> T (IN REF. 3).

SQ SEQUENCE 202 AA; 22494 MW; C0C853BE0032B541 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 202;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYI 119
|||
DB 13 ALEPYI 18

RESULT 106
NU6M_TRIRU STANDARD; PRT; 203 AA.
ID NU6M_TRIRU STANDARD; PRT; 203 AA.
AC Q36836;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
GN ND6 OR NADH6.
OS Trichophyton rubrum.
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Oryziales; Arthrodermataceae; mitosporic Arthrodermataceae;
CC Trichophyton.
OX NCBI_TaxID=5551;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IP 1817.89;
RX MEDLINE=96132111; PubMed=8593686;
RA de Bievre C., Dujon B.;
RT "Organisation of the mitochondrial genome of Trichophyton rubrum. DNA
RT sequence analysis of the ND4 gene, the ATPase subunit-6 gene, the
RT ribosomal RNA small-subunit gene, the ND6 gene, the COXII gene, the
RT ATPase subunit-8 gene and six tRNA genes that correspond respectively
RT to the tyrosine, lysine, glutamine, asparagine, isoleucine and
RT tryptophan isocodons.";
RL Curr. Genet. 28:553-559(1995).

CC -/- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
DR EMBL; X88896; CAA61356.1; -.
DR PIR; S65033; S65033.
DR InterPro; IPR001457; Oxidored g3.
DR Pfam; PF00499; Oxidored g3; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 203 AA; 22656 MW; 046DAB3716E02802 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 202 SYSFWL 207
DB 175 SYSFWL 180

RESULT 107
SOD2_PLEBO
ID SOD2_PLEBO STANDARD; PRT; 206 AA.
AC P50059;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn] 2 (EC 1.15.1.1).
GN SOD2.
OS Plectonema boryanum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
OX NCBI_TaxID=1184;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX 485;
RX MEDLINE=95164530; PubMed=7860607;
RA Campbell W.S., Laudenbach D.B.;
RT "Characterization of four superoxide dismutase genes from a filamentous cyanobacterium."
RL J. Bacteriol. 177:964-972 (1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- INDUCTION: BY METHYL VIOLOGEN, AND UNDER CONDITIONS OF IRON OR NITROGEN STRESS.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase family.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U17610; AAA69952.1; -.
CC HSP; P09214; INMG.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC PRODOM; PD000475; SODismutase; 1.
CC PROSITE; PS00089; SOD MN; 1.
KW Oxidoreductase; Metal-binding; Manganese; Multigene family.
FT METAL 27 27 MANGANESE (BY SIMILARITY).
FT METAL 82 82 MANGANESE (BY SIMILARITY).
FT METAL 165 165 MANGANESE (BY SIMILARITY).
FT METAL 169 169 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 206 AA; 23456 MW; B149F228DB49091E CRC64;

Query Match 2.5%; Score 6; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 202 SYSFWL 207
DB 175 SYSFWL 180

RESULT 108
JAG_BACHD
ID JAG_BACHD STANDARD; PRT; 207 AA.
AC Q95CA6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE JAG protein homolog.
GN JAG OR BH4063.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99356711; PubMed=10427704;
RA Takami H., Masui N., Nakasone K., Horikoshi K.;
RT "Replication origin region of the chromosome of alkaliphilic Bacillus halodurans C-125."
RL Biosci. Biotechnol. Biochem. 63:1134-1137 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058112;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N., Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S., Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331 (2000).
CC -!- SIMILARITY: Contains 1 KH domain.
CC -!- SIMILARITY: Contains 1 R3H domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AB013492; BAA82681.1; -.
CC EMBL; AP001520; BAB07782.1; -.
CC PIR; G84157; G84157.
CC InterPro; IPR001374; R3H.
CC Pfam; PF01424; R3H; 1.
CC SMART; SM00393; R3H; 1.
KW RNA-binding; Complete proteome.
FT DOMAIN 91 141 KH.
FT DOMAIN 150 202 R3H.
SQ SEQUENCE 207 AA; 23130 MW; BE02AF774460A632 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 59 QDLGTL 64
DB 98 QDLGTL 103

RESULT 109
NUGM PARTE
ID NUGM PARTE STANDARD; PRT; 209 AA.
AC P15600;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

DE NADH-ubiquinone oxidoreductase subunit 9 (EC 1.6.5.3) (EC 1.6.99.3)
DE (Protein P1).
GN NAD9.
OS Parametium tetraurelia.
OC Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Parametium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Stock 51;
RX MEDLINE=90174913; PubMed=2308823;
RA Pritchard A.E., Seilhamer J.J., Mahalingam R., Sable C.L.,
RA Venuti S.E., Cummings D.J.,
RT Nucleotide sequence of the mitochondrial genome of Parametium.;
RL Nucleic Acids Res. 18:173-180(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87055241; PubMed=3023187;
RA Pritchard A.E., Seilhamer J.J., Cummings D.J.,
RT "Parametium mitochondrial DNA sequences and RNA transcripts for
RT cytochrome oxidase subunit I, URFL, and three ORFs adjacent to the
RT replication origin."
RL Gene 44:243-253(1986).
CC -|- FUNCTION: Transfer of electrons from NADH to the respiratory
CC chain. The immediate electron acceptor for the enzyme is believed
CC to be ubiquinone (By similarity).
CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -|- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC -|- SUBUNIT: Complex I is composed of about 30 different subunits.
CC -|- SUBCELLULAR LOCATION: Matrix and cytoplasmic side of the
CC mitochondrial inner membrane.
CC -|- SIMILARITY: Belongs to the complex I 30 kDa subunit family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15917; CAA34058.1; -;
CC EMBL; M15275; AAA79259.1; -;
CC FIR; S07725; S07725.
CC InterPro; IPR001268; Complex1_30K.
CC Pfam; PF00329; complex1_30kd; 1.
CC ProDom; PD001581; Complex1_30K; 1.
CC PROSITE; PS00542; COMPLEX1_30K; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW SEQUENCE 209 AA; 23529 MW; D3D4477DC0BFB43 CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 169 GQALAS 174
DB 63 GQALAS 68
RESULT 110
GTS1 CAEEL STANDARD; PRT; 210 AA.
ID GTS1 CAEEL STANDARD; PRT; 210 AA.
AC Q9607;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable glutathione S-transferase R07B1.4 (EC 2.5.1.18)
DE (GST class-sigma).
DE R07B1.4.
GN R07B1.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: Conjugation of reduced glutathione to a wide number of
CC exogenous and endogenous hydrophobic electrophiles (By
CC similarity).
CC -|- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -|- SIMILARITY: Belongs to the GST superfamily. Sigma family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48621; CAA89541.2; -;
CC HSP; P24472; IGUK.
CC WormPep; R07B1.4; CE30562.
CC GO; GO:0004364; F:glutathione transferase activity; NAS.
CC GO; GO:0006803; P:glutathione conjugation reaction; NAS.
CC GO; GO:0004046; GST_Cterm.
CC InterPro; IPR004045; GST_Nterm.
CC InterPro; IPR004045; GST_C; 1.
CC Pfam; PF00043; GST_N; 1.
CC Pfam; PF02798; GST_N; 1.
KW Hypothetical protein; Transferase.
SQ SEQUENCE 210 AA; 23876 MW; 0E6010336B385D9B CRC64;
Query Match 2.5%; Score 6; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 26 SQTAL 31
DB 62 SQTAL 67
RESULT 111
RS3 OCEIH STANDARD; PRT; 213 AA.
ID RS3 OCEIH STANDARD; PRT; 213 AA.
AC P59182;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPSC OR OBO125.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -|- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -|- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -|- SIMILARITY: Belongs to the S3P family of ribosomal proteins.

```
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP004593; BAC12081.1; -
CC HAMAP; MF 01309; -; 1.
CC InterPro: IPR004087; KH dom.
CC InterPro: IPR009019; KH prok.
CC InterPro: IPR004044; KH TYPE 2.
CC InterPro: IPR001351; Ribosomal_S3_C.
CC InterPro: IPR008282; Ribosomal_S3_N.
CC InterPro: IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRfams; TIGR01009; rpsC_bact; 1.
CC PROSITE; PS00823; KH TYPE 2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
CC FT DOMAIN 38 106 KH TYPE-2.
CC SEQUENCE 213 AA; 23534 MW; 3543002B6C3B2934 CRC64;
CC -----
Query Match 2.5%; Score 6; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 EGTVP 41
DB 169 EGTVP 174
|||||
RESULT 112
SODF_AERPE STANDARD; PRT; 214 AA.
AC QY8H8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).
GN SOD OR APE0741.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OC NCBI_TaxID=56636;
OX [1]
RN SEQUENCE FROM N.A.
RA Yamano S.;
RT "A cambialistic SOD in a strictly aerobic hyperthermophilic archaeon,
RT Aeropyrum pernix."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kougai H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
```

```
CC -!- COFACTOR: Binds 1 manganese or iron ion per subunit (By
CC similarity). Belongs to the iron/manganese superoxide dismutase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB012621; BAA76442.1; -
CC EMBL: AP000060; BAA79718.1; -
CC PIR; F72664; F72664.
CC HSP; Q08713; I306.
CC InterPro: IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe C; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Iron; Complete proteome.
CC KW METAL 31 31 MANGANESE OR IRON (BY SIMILARITY).
CC FT METAL 79 79 MANGANESE OR IRON (BY SIMILARITY).
CC FT METAL 165 165 MANGANESE OR IRON (BY SIMILARITY).
CC FT METAL 169 169 MANGANESE OR IRON (BY SIMILARITY).
CC SEQUENCE 214 AA; 24577 MW; 641122779485DF0A CRC64;
CC -----
Query Match 2.5%; Score 6; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 114 ALRPI 119
DB 18 ALRPI 23
|||||
RESULT 113
FLAI_METJA STANDARD; PRT; 217 AA.
AC Q58301;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B1 precursor.
GN FLAB1 OR MJ0891.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96377999; PubMed=8689087;
RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Flagellin is the subunit protein which polymerizes to
CC form the filaments of archaeal flagella.
CC -!- SIMILARITY: Belongs to the archaeal flagellin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; U67533; AAB98894.1; -;
DR PIR; C64411; C64411.
DR TIGR; MJ0891; -;
DR InterPro; IPR002774; Arch_flagellin.
DR Pfam; PF01917; Arch_flagellin; 1.
KW Flagellum; Multigene family; Complete proteome.
FT PROPEP 1 12
FT CHAIN 13 217
SQ SEQUENCE 217 AA; 22700 MW; 4374437380061565 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKGRG 7
DB 7 LKGRG 12

RESULT 114

FLA2_METJA STANDARD; PRT; 217 AA.
AC Q58302;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B2 precursor.
GN FLA2 OR MJ0892.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
CC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervilave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).

CC -!- FUNCTION: Flagellin is the subunit protein which polymerizes to
CC form the filaments of archaeal flagella.

CC -!- SIMILARITY: Belongs to the archaeal flagellin family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; U67533; AAB98895.1; -;
DR PIR; D64411; D64411.
DR TIGR; MJ0892; -;
DR InterPro; IPR002774; Arch_flagellin.
DR Pfam; PF01917; Arch_flagellin; 1.
KW Flagellum; Multigene family; Complete proteome.
FT PROPEP 1 12
FT CHAIN 13 217
SQ SEQUENCE 217 AA; 22700 MW; 4374437380061565 CRC64;

SQ SEQUENCE 217 AA; 22577 MW; 5E9D9435C243A82D CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKGRG 7
DB 7 LKGRG 12

RESULT 115

RS3_BACST STANDARD; PRT; 217 AA.
AC P23309;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3 (B92) (BS3/B94).
GN R8SC
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCA 1503;
RX MEDLINE=91025633; PubMed=2222862;
RA Kroemer W.J., Hatakeyama T., Kimura M.;
RT "Nucleotide sequences of Bacillus stearothermophilus ribosomal
protein genes: part of the ribosomal S10 operon.";
RL Biol. Chem. Hoppe-Seyler 371:631-636(1990).
RN [2]

RP SEQUENCE OF 1-15.

RC STRAIN=10;
RX MEDLINE=75019590; PubMed=4607606;
RA Yaguchi M., Matheson A.T., Visentin L.P.;
RT "Prokaryotic ribosomal proteins: N-terminal sequence homologies and
structural correspondence of 30 S ribosomal proteins from Escherichia
coli and Bacillus stearothermophilus.";
RL FEBS Lett. 46:296-300(1974).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).

CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).

CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.

CC -!- SIMILARITY: Contains 1 KH type-2 domain.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; X54994; CAA38740.1; -;
DR PIR; S10613; S10613.

DR HAMAP; MF 01309; -; 1.

DR InterPro; IPR004087; KH dom.

DR InterPro; IPR009019; KH_prot.

DR InterPro; IPR004044; KH_TYPE_2.

DR InterPro; IPR001351; Ribosomal_S3_C.

DR InterPro; IPR008282; Ribosomal_S3_N.

DR InterPro; IPR005704; S3_bact.

DR Pfam; PF00013; KH; 1.

DR Pfam; PF00189; Ribosomal_S3_C; 1.

DR Pfam; PF00417; Ribosomal_S3_N; 1.

DR SMART; SMO0322; KH; 1.

DR TIGRFAWS; TIGR01009; rpsC_bact; 1.

DR PROSITE; PS00823; KH_TYPE_2; 1.

DR PROSITE; PS00548; RIBOSOMAL_S3; 1.

DR Ribosomal protein; RNA-binding; rRNA-binding.

```

FT INIT MET 0 0
FT DOMAIN 37 105 KH TYPE-2.
SQ SEQUENCE 217 AA; 24458 MW; 5A473C921758718C CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 168 EGTVP 173

RESULT 116
RS3_BACSU STANDARD; PRT; 217 AA.
AC P21465;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3 (BS) (BS2).
GN RPSC OR BSU01220.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG38;
RX MEDLINE=98037503; PubMed=9371452;
Li X., Lindahl L., Sha Y., Zeng J.M.;
RT "Analysis of the Bacillus subtilis S10 ribosomal protein gene cluster
RT identifies two promoters that may be responsible for transcription of
RT the entire 15-kilobase S10-spc-alpha cluster.";
RL J. Bacteriol. 179:7046-7054(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124188; PubMed=8969501;
Yasumoto K., Liu H., Jeong S.M., Ohashi Y., Kakinuma S.,
Tanaka K., Kawamura F., Yoshikawa H., Takahashi H.;
RT "Sequence analysis of a 50 kb region between spoOH and rrmH on the
RT Bacillus subtilis chromosome.";
RL Microbiology 142:3039-3046(1996).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferraci A., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haza K., Haech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue J.C., Medigue C.,
Medina N., Mollado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presceca E., Fujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
Viari A., Wambutt R., Wedler B., Wedler H., Weitzenecker T.,
```

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IL1403;
RM MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
in the 70S ribosome, positioning it for translation (By
similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006438; AK06191.1; -;
CC PIR; E86886; E86886.
CC HAMAP; MF_01309; -; 1.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR009019; KH_prok.
CC InterPro; IPR004044; KH_TYPE_2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC InterPro; IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRfams; TIGR01009; rpsC_bact; 1.
CC PROSITE; PS50823; KH_TYPE_2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 217 AA; 24034 MW; CF5937A341B76BD4 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVPL 41
DB 169 EGTVPL 174
|||||
RESULT 118
RS3_STAEP STANDARD; PRT; 217 AA.
AC Q99527;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SAV2244 OR SA2041 OR MW2163,
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=156878, 156879, 156620;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus";
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA";
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
in the 70S ribosome, positioning it for translation (By
similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003364; BAB58406.1; -;
CC EMBL; AP003366; BAB43336.1; -;
CC EMBL; AP004829; BAB96028.1; -;
CC PIR; G90021; G90021.
CC HAMAP; MF_01309; -; 1.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR009019; KH_prok.
CC InterPro; IPR004044; KH_TYPE_2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC InterPro; IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRfams; TIGR01009; rpsC_bact; 1.
CC PROSITE; PS50823; KH_TYPE_2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 217 AA; 24100 MW; 174CA582EB0DF917 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVPL 41
DB 169 EGTVPL 174
|||||
RESULT 119
RS3_STAEP STANDARD; PRT; 217 AA.
AC Q8CRG6;

```
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SE1818.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RA [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228.
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qiu Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,
RA "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).",
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE016750; AAC05459.1;
CC DR HAMAP: MF 01309; -; 1.
CC DR InterPro: IPR004087; KH_dom.
CC DR InterPro: IPR009019; KH_prok.
CC DR InterPro: IPR004044; KH_TYPE_2.
CC DR InterPro: IPR001351; Ribosomal_S3_C.
CC DR InterPro: IPR008282; Ribosomal_S3_N.
CC DR InterPro: IPR005704; S3_bact.
CC DR Pfam: PF00013; KH; 1.
CC DR Pfam: PF00189; Ribosomal_S3_C; 1.
CC DR Pfam: PF00417; Ribosomal_S3_N; 1.
CC DR SMART: SM00322; KH; 1.
CC DR TIGRFAMs: TIGR01009; rpsC_bact; 1.
CC DR PROSITE: PS00823; KH_TYPE_2; 1.
CC DR PROSITE: PS00548; RIBOSOMAL_S3; 1.
CC KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 217 AA; 24189 MW; 42488FE049F9E751 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred.No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 169 EGTVP 174
|||||
RESULT 120
RS3_STRM STANDARD; PRT; 217 AA.
AC P59186;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR RS3 OR SMU.2021.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SE1818.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RA [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228.
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qiu Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,
RA "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).",
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE015024; AAN59624.1; ALT_INIT.
CC DR HAMAP: MF 01309; -; 1.
CC DR InterPro: IPR004087; KH_dom.
CC DR InterPro: IPR009019; KH_prok.
CC DR InterPro: IPR004044; KH_TYPE_2.
CC DR InterPro: IPR001351; Ribosomal_S3_C.
CC DR InterPro: IPR008282; Ribosomal_S3_N.
CC DR InterPro: IPR005704; S3_bact.
CC DR Pfam: PF00013; KH; 1.
CC DR Pfam: PF00189; Ribosomal_S3_C; 1.
CC DR Pfam: PF00417; Ribosomal_S3_N; 1.
CC DR SMART: SM00322; KH; 1.
CC DR TIGRFAMs: TIGR01009; rpsC_bact; 1.
CC DR PROSITE: PS00823; KH_TYPE_2; 1.
CC DR PROSITE: PS00548; RIBOSOMAL_S3; 1.
CC KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 29 97 KH TYPE-2.
SQ SEQUENCE 217 AA; 24122 MW; CC8EB247CF331538 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred.No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 169 EGTVP 174
|||||
RESULT 121
RS3_STRPN STANDARD; PRT; 217 AA.
AC Q9WW37;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SP0215 OR SFR0195.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RA [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6, SP#5, and ZR1;
```


RX MEDLINE=20145398; PubMed=10681347;
RA Adrian P.V., Zhao W., Black T.A., Shaw K.J., Hare R.S., Klugman K.P.;
RT "Mutations in ribosomal protein L16 conferring reduced susceptibility
to evernimycin (SCH27899): implications for mechanism of action."
RL Antimicrob. Agents Chemother. 44:732-738 (2000).
RN [2]
RN
RP
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguino S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.;
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae."
RL Science 293:498-506 (2001).
RN [3]
RN
RP
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsumura P.,
RA McAnren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717 (2001).
CC
CC
CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
in the 70S ribosome, positioning it for translation (By
similarity).
CC
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
with proteins S10 and S14 (By similarity).
CC
CC -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC
CC -1- SIMILARITY: Contains 1 KH type-2 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AF126059; AAD33262.1; ALT_INIT.
CC EMBL; AF126060; AAD33271.1; ALT_INIT.
CC EMBL; AF126061; AAD33280.1; ALT_INIT.
CC EMBL; AE007335; AAK74395.1; -
CC EMBL; AE008402; AAX98999.1; -
CC EMBL; B95025; B95025.
CC EMBL; C97896; C97896.
CC TIGR; SP0215; -
CC HAVAP; MF 01309; -; 1.
CC InterPro; IPR004087; KH dom.
CC InterPro; IPR009019; KH prok.
CC InterPro; IPR004044; KH TYPE 2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC InterPro; IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRfam; TIGR01009; rpsC_bact; 1.
CC PROSITE; PS00823; KH TYPE 2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.

KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 217 AA; 24046 MW; B54CA0663A248663 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 EGVVPL 41
Db 169 EGVVPL 174
RESULT 122
RS3_STRPY
ID RS3_STRPY STANDARD; PRT: 217 AA.
AC Q9A1W8; P59185; Q8K8X2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SPY0056 OR SPYM3_0046 OR SPS0048 OR SPYM18_0057 OR GBS0064 OR
SAG0064.
OS Streptococcus Pyogenes.
OS Streptococcus Pyogenes (serotype M3).
OS Streptococcus Pyogenes (serotype M18).
OS Streptococcus agalactiae (serotype III), and
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 198466, 186103, 216495, 216466;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.pyogenes; STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.pyogenes; STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=2213808; PubMed=1212206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mamarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
phage-encoded toxins, the high-virulence phenotype, and clone
emergence."
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.pyogenes; STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=2268278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
large-scale genomic rearrangement in invasive strains and new insights
into phage evolution."
RL Genome Res. 13:1042-1055 (2003).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.pyogenes; STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gempel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18

```

QY      36 EGTVPVL 41
DB      169 EGTVPVL 174
|||||
|||||

RESULT 123
RS3_LISMO STANDARD; PRT; 218 AA.
ID RS3_LISMO AC Q927L3;
AC Q927L3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3
DE 30S ribosomal protein S3
GN RPSC OR LMO2626 OR LIN2775.
OS Listeria monocytogenes, and
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NCBI_TaxID=1639, 1642;
[1] SEQUENCE FROM N.A.
RN SPECIES=L.monocytogenes, and L.innocua;
RC STRAIN=SGD-e / Serovar 1/2a, and CLIP 11362 / Serovar 6a;
RC MEDLINE=21537279; PubMed=11679669;
RA Glaser F., Frangeul P., Blocher C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Blocher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fslhi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaert U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Medjari H.,
RA Nordieck G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,
RA Vazquez-Roland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852 (2001).
CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -1- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL591983; CAD00704.1; -
DR EMBL; AL596173; CAC98001.1; -
DR PIR; AB1403; AB1403.
DR PIR; A11778; A11778.
DR Listlist; LINQ2775; -.
DR Listlist; LMO02626; -.
DR HAMAP; MF_01309; -.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR009019; KH_prok.
DR InterPro; IPR004044; KH_TYPE_2.
DR InterPro; IPR001351; Ribosomal_S3_C.
DR InterPro; IPR008282; Ribosomal_S3_N.
DR InterPro; IPR005704; S3_bact.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF00189; Ribosomal_S3_C; 1.
DR Pfam; PF00417; Ribosomal_S3_N; 1.
DR SMART; SMO0322; KH; 1.
DR TIGRfams; TIGR01009; rpsC_bact; 1.
DR PROSITE; PSS0823; KH_TYPE_2; 1.
DR PROSITE; PSS0548; RIBOSOMAL_S3; 1.

```

DR PROSITE; PS00548; RIBOSOMAL_S3; 1
KW Ribosomal protein; rRNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 218 AA; 24542 MW; 4351A7E2CD75418A CRC64;
Query Match 2.5%; Score 6; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 EGTVPPL 41
DB 169 EGTVPPL 174

RESULT 124
RS3_BACAA
ID RS3_BACAA STANDARD; PRT; 219 AA.
AC Q81VS4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3.
GN RPSC OR BA0116.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=198094;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=22608414; PubMed=12721629;
RX Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holzapfel E.K., Ostad O.A., Heigason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Bentley J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nieman W.C.,
RA Hazen A., Cline R., Redmond C., Twaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3p family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE017024; AAP24170.1; -;
CC TIGR; BA0116; -;
CC HAMAP; MF_01309; -; 1.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR009019; KH_prot.
CC InterPro; IPR004044; KH_TYPE_2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC InterPro; IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRFRAMES; TIGR01009; rpsc_bact; 1.
CC PROSITE; PS50823; KH_TYPE_2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 219 AA; 24294 MW; C7BC458B0855AA1D CRC64;
Query Match 2.5%; Score 6; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 EGTVPPL 41
DB 169 EGTVPPL 174

RESULT 125
RS3_BACCE
ID RS3_BACCE STANDARD; PRT; 219 AA.
AC Q81J36;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3.
GN RPSC OR BC0137.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=226900;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=22608415; PubMed=12721630;
RX Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Fusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyripides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3p family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016998; AAP07218.1; -;
CC HAMAP; MF_01309; -; 1.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR009019; KH_prot.
CC InterPro; IPR004044; KH_TYPE_2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC InterPro; IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRFRAMES; TIGR01009; rpsc_bact; 1.
CC PROSITE; PS50823; KH_TYPE_2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 219 AA; 24294 MW; C7BC458B0855AA1D CRC64;
Query Match 2.5%; Score 6; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 EGTVPPL 41
DB 169 EGTVPPL 174


```
Db 37 ALEPYI 42
|||||
Query Match 2.5%; Score 6; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LGTLGS 66
DB 10 LGTLGS 15

RESULT 128
PGC2 HUMAN STANDARD; PRT; 223 AA.
AC O15173;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Membrane associated progesterone receptor component 2 (Progesterone
DE membrane binding protein) (Steroid receptor protein DG6).
GN PORWC2 OR PWPB OR DG6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98368553; PubMed=9705155;
RA Gerdes D., Wehling M., Leube B., Falkenstein E.;
RT "Cloning and tissue expression of two putative steroid membrane
RT receptors.";
RL Biol. Chem. 379:907-911(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Parmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Basak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: IS A RECEPTOR FOR STEROIDS (POTENTIAL).
CC -1- SIMILARITY: Belongs to the MAPR family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ002030; CAA05152.1; -
DR EMBL; BC016892; AAH16892.1; -
DR Genew; HGNC:16089; PGRWC2.
DR MIM; 607735; -
DR GO; GO:0016021; C: integral to membrane; TAS.
DR GO; GO:0005496; F: steroid binding; TAS.
DR GO; GO:0003707; F: steroid hormone receptor activity; TAS.
DR InterPro; IPR001199; Cyt_B5.
DR Pfam; PF00173; heme_1; 1.
KW Receptor; Steroid-binding; Transmembrane; Microsome.
FT TRANSMEM 42
FT POTENTIAL 66
SQ SEQUENCE 223 AA; 23818 MW; BE36229ED0FF3AD CRC64;
```

```
Query Match 2.5%; Score 6; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 LEKIIS 234
DB 171 LEKIIS 176

RESULT 130
LIPB MYCLE STANDARD; PRT; 235 AA.
ID LIPB MYCLE
```

```
Query Match 2.5%; Score 6; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 LEKIIS 234
DB 171 LEKIIS 176

RESULT 130
LIPB MYCLE STANDARD; PRT; 235 AA.
ID LIPB MYCLE
```

```
Query Match 2.5%; Score 6; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 LEKIIS 234
DB 171 LEKIIS 176

RESULT 130
LIPB MYCLE STANDARD; PRT; 235 AA.
ID LIPB MYCLE
```

AC 032961;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lipote-protein ligase B (EC 6.-.-.-) (Lipote biosynthesis protein
B).
DE
GN LIPB OR MLC0859 OR MLCB22.19.
OS Mycobacterium leprae.
CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher T., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Involved in the attachment of lipoyl groups to proteins,
CC by creating an amide linkage that joins the free carboxyl group of
CC lipoyl acid to the epsilon-amino group of a specific lysine
CC residue in lipoylated proteins (By similarity).
CC -!- PATHWAY: Lipote biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the lipB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z98741; CAB11384.1; -
CC EMBL; AL583920; CAC31240.1; -
CC PIR; T44894; T44894.
CC Lepronia; MLC0859; -
CC HAMAP; MF_00013; -; 1.
CC InterPro; IPR004143; BPL LipA LipB.
CC InterPro; IPR000544; Lipotease B.
CC Pfam; PF03099; BPL LipA LipB; 1.
CC ProDom; PD006086; Lipotease; 1.
CC TIGRFAMs; TIGR00214; lipB; 1.
CC PROSITE; PS01313; LIPB; 1.
CC Ligase; Complete proteome.
KW SEQUENCE 235 AA; 25002 MW; 353FA05D2E4A4A67D CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 96 WLSTPA 101
DB 224 WLSTPA 229
RESULT 131
SOD1_PLEBO
ID SOD1_PLEBO STANDARD; PRT; 248 AA.
AC P50058;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn] 1 precursor (EC 1.15.1.1).

GN SOD1.
OS Plectonema boryanum.
CC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
OX NCBI_TaxID=1184;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UTEX 485;
RX MEDLINE=95164530; PubMed=7860607;
RA Campbell W.S., Laudenbach D.E.;
RT "Characterization of four superoxide dismutase genes from a
RT filamentous cyanobacterium."
RL J. Bacteriol. 177:964-972(1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- INDUCTION: Constitutively expressed.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U17609; AAA6950.1; -
CC HSP; P00448; IVEW.
CC InterPro; IPR000437; Prok lipoprot_S.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe C; 1.
CC PRINTS; PR01703; MNSODISMTASE.
CC ProDom; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Multigene family; Signal.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 248 SUPEROXIDE DISMUTASE [MN] 1.
FT METAL 68 68 MANGANESE (BY SIMILARITY).
FT METAL 123 123 MANGANESE (BY SIMILARITY).
FT METAL 208 208 MANGANESE (BY SIMILARITY).
FT METAL 212 212 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 248 AA; 27955 MW; 85697F08623EAD1 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 248;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 114 ALEPYI 119
DB 55 ALEPYI 60
RESULT 132
TRYP_PLEPL
ID TRYP_PLEPL STANDARD; PRT; 250 AA.
AC P35034;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Pleuronectes platessa (Plaice)
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
CC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
CC Pleuronectoidae; Pleuronectidae; Pleuronectes.
OX NCBI_TaxID=8262;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;

```
RA Leaver M.J., George S.G.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X56744; CA440068.1; -
DR F01; S31384; S31384.
DR HSP; P00761; 1EPT.
DR MEROPS; S01151; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR01254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 22
FT CHAIN 23 250
FT ACT_SITE 62 62
FT ACT_SITE 106 106
FT ACT_SITE 203 203
FT DISULFID 29 163
FT DISULFID 47 63
FT DISULFID 133 236
FT DISULFID 140 209
FT DISULFID 174 198
FT DISULFID 199 223
FT SITE 197 197
FT SITE 197 197
SQ SEQUENCE 250 AA; 27527 MW; 637D96185CIABAA CRC64;

Query Match 2.5%; Score 6; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
DB 201 GDSGSP 206

RESULT 133
ID_GRAA_MOUSE STANDARD; PRT; 260 AA.
AC P11032; P15118;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Granzyme A precursor (EC 3.4.21.78) (T cell-specific serine protease
DE 1) (TSP-1) (CTLA-3) (Fragmentin 1) (Autocrine thymic lymphoma
DE granzyme-like serine protease).
GN GZMA OR CTLA3 OR CTLA-3 OR MTSP-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=92347841; PubMed=1639378;
RA Ebnet K., Kramer M.D., Simon M.M.;
RA "Organization of the gene encoding the mouse T-cell-specific serine
RT proteinase 'granzyme A'";
Genomics 13:502-508(1992).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RX MEDLINE=89144553; PubMed=2976140;
RA Bogenberger J., Haas M.;
RA "cDNA clones from autocrine thymic lymphoma cells encode two
RT mitogenic proteins, a serine protease and a truncated T-cell receptor
RT beta-chain.";
Oncogene Res. 3:301-312(1988).
[3]
RP SEQUENCE FROM N.A. (ISOFORMS HF1 AND HF2).
RX MEDLINE=93094270; PubMed=1460043;
RA Hershberger R.J., Gershenfeld H.K., Weissman I.L., Su L.;
RA "Genomic organization of the mouse granzyme A gene. Two mRNAs encode
RT the same mature granzyme A with different leader peptides.";
J. Biol. Chem. 267:25488-25493(1992).
[4]
RP SEQUENCE OF 12-260 FROM N.A.
RX MEDLINE=86208119; PubMed=2422755;
RA Gershenfeld H.K., Weissman I.L.;
RA "Cloning of a cDNA for a T cell-specific serine protease from a
RT cytotoxic T lymphocyte.";
Science 232:854-858(1986).
[5]
RP SEQUENCE OF 29-48.
RX MEDLINE=87215912; PubMed=3555842;
RA Masson D., Tschopp J.;
RA "A family of serine esterases in lytic granules of cytolytic T
RT lymphocytes.";
Cell 49:679-685(1987).
[6]
RP SEQUENCE OF 29-53.
RX MEDLINE=87030960; PubMed=3533635;
RA Masson D., Zamaï M., Tschopp J.;
RA "Identification of granzyme A isolated from cytotoxic T-lymphocyte-
RT granules as one of the proteases encoded by CTL-specific genes.";
FEBS Lett. 208:84-88(1986).
[7]
RP SEQUENCE OF 29-46 FROM N.A.
RX MEDLINE=88255076; PubMed=3260181;
RA Simon H.G., Fruth U., Eckerskorn C., Lottspeich F., Kramer M.D.,
RA Nerz G., Simon M.M.;
RA "Induction of T cell serine proteinase 1 (TSP-1)-specific mRNA in
RT mouse T lymphocytes.";
Eur. J. Immunol. 18:855-861(1988).
[8]
RP SEQUENCE OF 1-37 FROM N.A.
RX MEDLINE=88272336; PubMed=3292396;
RA Jenne D.E., Tschopp J.;
RA "Granzymes, a family of serine proteases released from granules of
RT cytolytic T lymphocytes upon T cell receptor stimulation.";
Immunol. Rev. 103:53-71(1988).
CC -!- FUNCTION: This enzyme is necessary for target cell lysis in cell-
CC mediated immune responses. It cleaves after Lys or Arg. May be
CC involved in apoptosis.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin,
CC type IV collagen and nucleolin. Preferential cleavage: Arg-|-Xaa,
CC Lys-|-Xaa >> Phe-|-Xaa in small molecule substrates.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Secreted; cytoplasmic granules of cytolytic
CC T-lymphocytes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=HF1;
CC IsoId=P11032-1; Sequence=Displayed;
CC Note=More abundant in lymphoid tissues than isoform HF2;
CC Name=HF2;
CC IsoId=P11032-2; Sequence=VSP_005377;
CC -!- TISSUE SPECIFICITY: Found in cytotoxic lymphocytes and in normal
CC lymphoid tissues such as thymus and spleen.
CC -!- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.
```


RP 3D-STRUCTURE MODELING.
 RX MEDLINE=9184501; PubMed=3237717;
 RA Murphy M.E.P., Moulton J., Bleackley R.C., Gershenfeld H.,
 RA Weissman I.L., James M.N.G.;
 RT "Comparative molecular model building of two serine proteinases from
 RL cytotoxic T lymphocytes.";
 RL Proteins 4:190-204(1988).
 CC !- FUNCTION: This enzyme is necessary for target cell lysis in cell-
 CC mediated immune responses. It cleaves after Lys or Arg. May be
 CC involved in apoptosis.
 CC !- CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin,
 CC type IV collagen and nucleolin. Preferential cleavage: Arg-|-Xaa,
 CC Lys-|-Xaa >> Phe-|-Xaa in small molecule substrates.
 CC !- SUBUNIT: Homodimer; disulfide-linked.
 CC !- SUBCELLULAR LOCATION: Secreted; cytoplasmic granules.
 CC !- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M18737; AAA52647.1; -.
 DR EMBL; BC015739; AAH15739.1; -.
 DR EMBL; U40006; AAD00009.1; -.
 DR PIR; A31372; A31372
 DR PDB; 1HF1; 15-OCT-94.
 DR MEROPS; S01.135; -.
 DR Genew; HGNC:4708; GZMA.
 DR MIM; 140050; -.
 DR GO; GO:0004277; F:Granzyme A activity; TAS.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001354; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0722; CHYNOTRYPsin.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolyase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
 KW Apoptosis; 3D-structure.
 FT SIGNAL 1 26
 FT PROPEP 27 28 ACTIVATION PEPTIDE.
 FT CHAIN 29 262 GRANZYME A.
 FT ACT_SITE 69 69 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 114 114 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 54 70 BY SIMILARITY.
 FT DISULFID 148 218 BY SIMILARITY.
 FT DISULFID 179 197 BY SIMILARITY.
 FT DISULFID 208 234 BY SIMILARITY.
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 30 30
 FT STRAND 33 34
 FT TURN 37 38
 FT TURN 41 42
 FT STRAND 43 48
 FT TURN 49 51
 FT STRAND 52 60
 FT TURN 61 62
 FT STRAND 63 66
 FT TURN 68 69
 FT STRAND 76 80
 FT STRAND 84 84
 FT TURN 85 86
 FT TURN 90 91
 FT STRAND 93 102
 FT TURN 104 105
 FT HELIX 108 110

FT TURN 113
 FT STRAND 116
 FT STRAND 120
 FT TURN 127
 FT TURN 129
 FT STRAND 130
 FT STRAND 134
 FT TURN 138
 FT TURN 144
 FT STRAND 147
 FT STRAND 155
 FT TURN 158
 FT STRAND 160
 FT STRAND 165
 FT STRAND 167
 FT HELIX 176
 FT TURN 181
 FT TURN 184
 FT TURN 193
 FT STRAND 195
 FT TURN 201
 FT STRAND 206
 FT TURN 209
 FT TURN 212
 FT STRAND 215
 FT TURN 219
 FT STRAND 221
 FT TURN 231
 FT TURN 234
 FT TURN 237
 FT STRAND 241
 FT TURN 246
 FT HELIX 252
 FT SEQUENCE 262 AA; 28966 MW; DA87363A0D92BAF4 CRC64;
 SQ
 Query Match 2.5%; Score 6; DB 1; Length 262;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GDSGSP 12
 DB 210 GDSGSP 215
 RESULT 135
 ELNE HUMAN STANDARD; PRT; 267 AA.
 AC P08246; P09649;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Leukocyte elastase precursor (SC 3.4.21.37) (Neutrophil elastase)
 DE (PMN elastase) (Bone marrow serine protease) (Medullasin).
 GN ELA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89374820; PubMed=2775493;
 RA Farley D., Travis J., Salvesen G.;
 RT "The human neutrophil elastase gene. Analysis of the nucleotide
 RL sequence reveals three distinct classes of repetitive DNA.";
 RL Biol. Chem. Hoppe-Seyler 370:737-744(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88067782; PubMed=3479752;
 RA Nakamura H., Okano K., Aoki Y., Shimizu H., Naruto M.;
 RT "Nucleotide sequence of human bone marrow serine protease
 RL (medullasin) gene.";
 RL Nucleic Acids Res. 15:9601-9601(1987).
 RN [3]
 RP SEQUENCE FROM N.A.

RA MEDLINE=89008342; PubMed=2302087;
RA Takahashi H., Nukiwa T., Yoshimura K., Quick C.D., States D.J.,
RA Holmes M.D., Whang-Peng J., Knutsen T., Crystal R.G.;
RT "Structure of the human neutrophil elastase gene.";
RL J. Biol. Chem. 263:14739-14747(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=90211319; PubMed=2322278;
RX Okano K., Aoki Y., Shimizu H., Naruto M.;
RA "Functional expression of human leukocyte elastase (HLE)/medullasin
RT in eukaryotic cells.";
RL Biochem. Biophys. Res. Commun. 167:1326-1332(1990).
RN [5]
RP SEQUENCE OF 30-267 FROM N.A.
RX MEDLINE=89032918; PubMed=2822677;
RX Okano K., Aoki Y., Sakurai T., Kajitani M., Kanai S., Shimazu T.,
RA Shimizu H., Naruto M.;
RT "Molecular cloning of complementary DNA for human medullasin: an
RT inflammatory serine protease in bone marrow cells.";
RL J. Biochem. 102:13-16(1987).
RN [6]
RP SEQUENCE OF 75-267 FROM N.A.
RX MEDLINE=88115408; PubMed=3422232;
RX Takahashi H., Nukiwa T., Bassett P., Crystal R.G.;
RA "Myelomonocytic cell lineage expression of the neutrophil elastase
RT gene.";
RL J. Biol. Chem. 263:2543-2547(1988).
RN [7]
RP SEQUENCE OF 30-247
RX MEDLINE=87175847; PubMed=3350808;
RX Sinha S., Watorek W., Karr S., Giles J., Bode W., Travis J.;
RA "Primary structure of human neutrophil elastase.";
RT Proc. Natl. Acad. Sci. U.S.A. 84:2228-2232(1987).
RN [8]
RP SEQUENCE OF 262-267.
RX MEDLINE=91315473; PubMed=1859409;
RX Aoki Y., Hase T.;
RT "The primary structure and elastolytic activity of medullasin (a
RT serine protease of bone marrow).";
RL Biochem. Biophys. Res. Commun. 178:501-506(1991).
RN [9]
RP PRELIMINARY SEQUENCE OF 30-103
RA Travis J., Giles J., Porcelli L., Reilly C.F., Baugh R., Powers J.;
RA (in) Protein degradation in health and disease, Ciba Foundation
RL Symposium, pp.75:51-68, Excerpta Medica, Amsterdam and Oxford (1980).
RN [10]
RP SEQUENCE OF 30-49.
RX MEDLINE=89315847; PubMed=2501794;
RX Gabay J.E., Scott R.W., Campanelli D., Griffith J., Wilde C.,
RA Marra M.N., Seeger M., Nathan C.F.;
RT "Antibiotic proteins of human polymorphonuclear leukocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5610-5614(1989).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (1.84 ANGSTROMS).
RX MEDLINE=89098932; PubMed=2911584;
RX Navia M.A., McKeever B.M., Springer J.P., Lin T.-Y., Williams H.R.,
RA Fluder E.M., Dorn C.P., Hoogsteen K.;
RT "Structure of human neutrophil elastase in complex with a peptide
RT chloromethyl ketone inhibitor at 1.84-A resolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7-11(1989).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=88271660; PubMed=3391280;
RX Wei A.-Z., Mayr I., Bode W.;
RA "The refined 2.3-A crystal structure of human leukocyte elastase in a
RT complex with a valine chloromethyl ketone inhibitor.";
RL FEBS Lett. 234:367-373(1988).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=87053808; PubMed=3640709;
RX Bode W., Wei A.-Z., Huber R., Meyer E., Travis J., Neumann S.;
RA "X-ray crystal structure of the complex of human leukocyte elastase
RT (PMN elastase) and the third domain of the turkey ovomucoid

RT inhibitor.";
RL EMBO J. 5:2453-2458(1986).
RN [14]
RP VARIANTS CH VAL-32; PHE-177 AND GLN-191.
RX MEDLINE=20047772; PubMed=10581030;
RX Horwitz M., Benson K.F., Person R.E., Aprikyan A.G., Dale D.C.;
RA "Mutations in ELA2, encoding neutrophil elastase, define a 21-day
RT biological clock in cyclic haematopoiesis.";
RL Nat. Genet. 23:433-436(1999).
CC -|- FUNCTION: Medullasin modifies the functions of natural killer
CC cells, monocytes and granulocytes.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin.
CC Preferential cleavage: Val-|-Xaa > Ala-|-Xaa.
CC -|- TISSUE SPECIFICITY: Bone marrow cells.
CC (CYCLIC NEUTROPENIA); AN AUTOSOMAL DOMINANT DISEASE IN WHICH
CC BLOOD-CELL PRODUCTION FROM THE BONE MARROW OSCILLATES WITH 21-DAY
CC PERIODICITY. CIRCULATING NEUTROPHILS VARY BETWEEN ALMOST NORMAL
CC NUMBERS AND ZERO. DURING INTERVALS OF NEUTROPENIA, AFFECTED
CC INDIVIDUALS ARE AT RISK FOR OPPORTUNISTIC INFECTION. MONOCYTES,
CC PLATELETS, LYMPHOCYTES AND RETICULOCYTES ALSO CYCLE WITH THE SAME
CC FREQUENCY.
CC -|- SIMILARITY: Belongs to peptidase family S1. Elastase subfamily.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03545; AAA52378.1; -
DR EMBL; Y00477; CAA68537.1; -
DR EMBL; X05875; CAA29299.1; -
DR EMBL; X05875; CAA29300.1; ALT_INIT.
DR EMBL; M20203; AAA36389.1; -
DR EMBL; M20199; AAA36389.1; JOINED.
DR EMBL; M20200; AAA36389.1; JOINED.
DR EMBL; M20201; AAA36389.1; JOINED.
DR EMBL; M34379; AAA36173.1; -
DR EMBL; D00187; BAA00128.1; -
DR PIR; A31976; ELHUL.
DR PDB; 1HNE; 15-OCT-89.
DR PDB; 1PPP; 31-JAN-94.
DR PDB; 1PPG; 31-JAN-94.
DR PDB; 1B0F; 18-NOV-98.
DR MEROPS; S01.131; -
DR Genew; HGNC:3309; ELA2.
DR MIM; 130130; -
DR MIM; 152800; -
DR GO; GO:0004234; F:macrophage elastase activity; TAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Glycoprotein; Signal; 3D-structure;
KW Disease mutation.
KW SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 29
FT CHAIN 30 267 LEUKOCYTE ELASTASE.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM.
FT ACT_SITE 117 117 CHARGE RELAY SYSTEM.
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM.
FT DISULFID 55 71
FT DISULFID 151 208
FT DISULFID 181 187
FT DISULFID 198 223

```

FT CARBOHYD 88      88      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124     124     N-LINKED (GLCNAC. . .)
FT CARBOHYD 173     173     N-LINKED (GLCNAC. . .)
FT VARIANT 32       32       G -> V (IN CH)
FT VARIANT 177     177     /FTId=VAR_009538.
FT VARIANT 191     191     V -> F (IN CH)
FT VARIANT 191     191     /FTId=VAR_009539.
FT CONFLICT 107    107     R -> Q (IN CH)
FT STRAND 31       31       /FTId=VAR_009540.
FT STRAND 34       34       N -> D (IN REF. 6).
FT TURN 38        38
FT TURN 39        39
FT TURN 42        42
FT TURN 43        43
FT TURN 44        44
FT TURN 50        50
FT TURN 52        52
FT TURN 62        62
FT TURN 63        63
FT STRAND 64      64
FT STRAND 66      66
FT HELIX 69       69
FT TURN 72       72
FT HELIX 77       77
FT STRAND 81      81
FT STRAND 88      88
FT TURN 89       89
FT TURN 94       94

Query Match      2.5%; Score 6; DB 1; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
Db 200 GDSGSP 205

RESULT 136
ZUPT_OCEIH
ID ZUPT_OCEIH STANDARD; PRT; 268 AA.
AC Q8ENQ1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc transporter zupt.
GN ZUPT OR OB2427.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=HTEB31 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments."
RT Nucleic Acids Res. 30:3927-3935 (2002).
RL NCBI_TaxID=182710;
CC -1- FUNCTION: Mediates zinc uptake. May also transport other divalent cations (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family. Zupt subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AP004601; BAC14383.1; -
CC HAMAP; MF_00548; -; 1.
DR

```

```

DR InterPro; IPR003689; Zn_transpt_Zip.
DR Pfam; PF02535; Zip; 1.
KW Transport; Zinc transport; Transmembrane; Complete proteome.
FT TRANSMEM 4       26     Potential.
FT TRANSMEM 35      57     Potential.
FT TRANSMEM 72      94     Potential.
FT TRANSMEM 126     145     Potential.
FT TRANSMEM 155     177     Potential.
FT TRANSMEM 189     211     Potential.
FT TRANSMEM 215     237     Potential.
FT TRANSMEM 250     267     Potential.
SQ SEQUENCE 268 AA; 29073 MW; 393A3C339277E0C9 CRC64;

Query Match      2.5%; Score 6; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
Db 156 AIAIAV 161

RESULT 137
YDGD_ECOLI
ID YDGD_ECOLI STANDARD; PRT; 273 AA.
AC P76176;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative protease ydgd precursor (EC 3.4.21.-).
GN YDGD OR B1598.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474 (1997).
RL NCBI_TaxID=562;
CC -1- SIMILARITY: Belongs to peptidase family S2B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000255; AAC74670.1; -
CC PIR; H64915; H64915.
CC MEROPS; S01.260; -.
CC Ecogen; EG3925; ydgd.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR008256; Peptidase_S1B_V8.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00839; V8PROTEASE.
CC SMART; SM00202; TrypSPC; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; Hydrolase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1         21     POTENTIAL.
FT CHAIN 22        273     PUTATIVE PROTEASE YDGD.
FT ACT_SITE 84      84     CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 223     223     CHARGE RELAY SYSTEM (BY SIMILARITY).
DR

```

```

SQ SEQUENCE 273 AA; 29277 NW; C8FCD018A59DBC62 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
Db 221 GDSGSP 226

RESULT 138
UPK_STRPN
ID UPK_STRPN STANDARD; PRT; 281 AA.
AC Q973C8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative undecaprenol kinase (EC 2.7.1.66) (Bacitracin resistance
protein).
GN UPK OR BACA OR SP0457 OR SPR0413.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heideberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Whitt M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., Gwin O., Salzborg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angioli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.";
RL Science 293:498-506(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=1154234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAlren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris P.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glaes J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
RN [3]
RP FUNCTION
RC STRAIN=0100993 / NCIMS 40794 / Serotype 3;
RX MEDLINE=20340958; PubMed=10878119;
RA Chalker A.F., Ingraham K.A., Lunsford R.D., Bryant A.P., Bryant J.,
RA Wallis N.G., Brokney J.P., Pearson S.C., Holmes D.J.;
RT "The bcaA gene, which determines bacitracin susceptibility in
Streptococcus pneumoniae and Staphylococcus aureus, is also required
for virulence.";
RL Microbiology 146:1547-1553(2000).
CC -!- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl
phosphate. Confers resistance to bacitracin. Is also required for
virulence.
CC -!- CATALYTIC ACTIVITY: ATP + undecaprenol = ADP + undecaprenyl
phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition
of peptidoglycan synthesis by sequestering undecaprenyl
diphosphate reducing the pool of lipid carrier available.
-!- SIMILARITY: Belongs to the upk family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; AE007357; AAK74617.1; -.
CC EMBL; AE008421; AAK99217.1; -.
DR PIR; E97923; E97923.
DR PIR; H95052; H95052.
DR TIGR; SP0457; -.
DR HAMAP; MF_01006; -.
DR InterPro; IPR003824; BACA.
DR Pfam; PF02673; BACA; 1.
KW Transferase; Kinase; Antibiotic resistance; Transmembrane;
Complete proteome.
FT TRANSMEM 10 32 POTENTIAL.
FT TRANSMEM 45 67 POTENTIAL.
FT TRANSMEM 111 133 POTENTIAL.
FT TRANSMEM 154 176 POTENTIAL.
FT TRANSMEM 191 213 POTENTIAL.
FT TRANSMEM 226 248 POTENTIAL.
FT TRANSMEM 263 280 POTENTIAL.
SQ SEQUENCE 281 AA; 31810 MW; 0575353B1BDF97C2 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 VKAGEL 229
Db 216 VKAGEL 221

RESULT 139
AQP6 HUMAN
ID AQP6 HUMAN STANDARD; PRT; 282 AA.
AC Q13520;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aquaporin 6 (Aquaporin-2 like) (hKID).
GN AQP6 OR AQP2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97001157; PubMed=8812490;
RA Ma T., Yang B., Kuo W.L., Verkman A.S.;
RT "cDNA cloning and gene structure of a novel water channel expressed
exclusively in human kidney: evidence for a gene cluster of
aquaporins at chromosome locus 12q13.";
RL Genomics 35:543-550(1996).
CC -!- FUNCTION: Forms a water-specific channel (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- DOMAIN: Aquaporins contain two tandem repeats each containing
three membrane-spanning domains and a pore-forming loop with the
signature motif Asn-Pro-Ala (NPA).
CC -!- SIMILARITY: Belongs to the MIP/aquaporin (TC 1.A.8) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; U48408; AAB41566.1; -.
CC DR HSP; P29972; IPOV.
CC DR Genew; HGNC:639; AQP6.
CC DR MIM; 601383; -.
CC DR GO; GO:0005887; C: integral to plasma membrane; TAS.
CC DR GO; GO:0015250; F: water channel activity; TAS.
CC DR GO; GO:0007588; P: excretion; TAS.
CC DR GO; GO:0006810; P: transport; TAS.
CC DR InterPro; IPR000425; MIP.
CC DR Pfam; PF00230; MIP; 1.
CC DR PRINTS; PR00783; MINTRINSIC.
CC DR ProDom; PD000295; MIP family; 1.
CC DR TIGRFAMs; TIGR00861; MIP; 1.
CC DR PROSITE; PS00221; MIP; 1.
CC KW Transport; Repeat; Transmembrane.
CC FT DOMAIN 1 30 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 31 48 POTENTIAL.
CC FT DOMAIN 49 54 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 55 73 POTENTIAL.
CC FT DOMAIN 74 99 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 100 121 POTENTIAL.
CC FT DOMAIN 122 141 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 142 162 POTENTIAL.
CC FT DOMAIN 163 168 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 169 188 POTENTIAL.
CC FT DOMAIN 189 214 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 215 236 POTENTIAL.
CC FT DOMAIN 237 282 CYTOPLASMIC (POTENTIAL).
CC FT SITE 82 84 NPA 1.
CC FT SITE 196 198 NPA 2.
CC SQ SEQUENCE 282 AA; 29562 MW; 990BECB5D9311ADD3 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPAT 14
Db 169 SGSPAT 174

RESULT 140
ID TRUB_FUSNN STANDARD; PRT; 287 AA.
AC QRSY8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE synthase) (Pseu synthase) (Pseudouridylate synthase) (Uracil
DE hydrolyase).
GN TRUB OR FN0635.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21886394; PubMed=11899109;
RA Kapatal V., Anderson I., Ivanova N., Resnik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyriades N., Overbeek R.,
RT Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25866.;
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC uracil-55 in the psi GC loop of transfer RNAs (By similarity).
CC -!- CATALYTIC ACTIVITY: uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H2O.
CC -!- SIMILARITY: Belongs to the pseudouridine synthase truB family.
CC Subfamily 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AB010575; AAL94831.1; -.
CC HAMAP; MF_01080; -.
CC InterPro; IPR004510; TruB_synth_N.
CC InterPro; IPR002501; TruB_N; 1.
CC Pfam; PF01509; TruB_N; 1.
CC TIGRFAMs; TIGR00431; TruB; 1.
KW tRNA processing; Lyase; Complete proteome.
FT ACT SITE 38 38 BY SIMILARITY.
SQ SEQUENCE 287 AA; 33092 MW; 5646C730F7BD6E23 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 287;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GQDLGT 63
Db 178 GQDLGT 183

RESULT 141
ID ZUPT_CAMJE STANDARD; PRT; 291 AA.
AC QPIN2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc transporter zupt.
DE Zinc transporter zupt.
GN ZUPT OR CJ0263.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RC MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Randle M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.;
RL Nature 433:665-668(2000).
CC -!- FUNCTION: Mediates zinc uptake. May also transport other divalent
CC cations (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family. Zupt
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AL139074; CAB72731.1; -.
CC -----

```
DR PIR, F81444; F81444.
DR HAMAP; MF 00548; -; 1.
DR InterPro; IPR003689; Zn_transpt_zip.
DR Pfam; PF02535; Zip; 1.
KW Transport; Zinc transport; Transmembrane; Complete proteome.
FT TRANSMEM 8
FT TRANSMEM 39
FT TRANSMEM 59
FT TRANSMEM 74
FT TRANSMEM 94
FT TRANSMEM 147
FT TRANSMEM 174
FT TRANSMEM 209
FT TRANSMEM 229
FT TRANSMEM 233
FT TRANSMEM 271
FT TRANSMEM 291
SQ SEQUENCE 291 AA; 31462 MW; 35A0E51E408E1CF2 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
DB 178 AIAIAV 183

RESULT 142
EFTS_XANAC STANDARD; PRT; 292 AA.
AC Q8PAV3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Elongation factor Ts (EF-Ts)
GN TSF OR XAC1421.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the EF-Ts family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AE012237; AAM40672.1; ALT INIT.
```

```

DR HAMAP; MF_00050; -; 1.
DR InterPro; IPR001816; EF TS.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00889; EF TS; 1.
DR Pfam; PF00627; UBA; 1.
DR TIGRfams; TIGR00116; tsf; 1.
DR PROSITE; PS01126; EF TS_1; 1.
DR PROSITE; PS01127; EF TS_2; 1.
KW Elongation factor; Protein biosynthesis; Complete proteome.
FT SITE 79 INVOLVED IN MG++ ION DISLOCATION FROM EF-
FT SITE 82 TU (BY SIMILARITY).
SQ SEQUENCE 292 AA; 31007 MW; A167B3F0AB576766 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 292;
Best Local Similarity 100.0%; Pred.No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIS 234
DB 221 LEKIIS 226
|||||
|||||

RESULT 144
EFTS_XYLFA STANDARD; PRT; 292 AA.
AC Q9PAD9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Elongation factor Ts (EF-Ts).
TSF OR XF2579.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.N.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.D., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.N., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.F., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silveira M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Teuhako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the EF-Ts family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004065; AAF95376.1; ALT_INIT.
CC HSSP; P02997; LEFU.
CC HAMAP; MF_00050; -; 1.
CC InterPro; IPR001816; UBA_domain.
CC InterPro; IPR000449; UBA_domain.
CC Pfam; PF00889; EF TS; 1.
CC Pfam; PF00627; UBA; 1.
CC TIGRfams; TIGR00116; tsf; 1.
CC PROSITE; PS01126; EF TS_1; 1.
CC PROSITE; PS01127; EF TS_2; 1.
KW Elongation factor; Protein biosynthesis; Complete proteome.
FT SITE 79 INVOLVED IN MG++ ION DISLOCATION FROM EF-
FT SITE 82 TU (BY SIMILARITY).
SQ SEQUENCE 292 AA; 31235 MW; 433061318FACB8E0 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 292;
Best Local Similarity 100.0%; Pred.No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIS 234
DB 221 LEKIIS 226
|||||
|||||

RESULT 145
EFTS_XYLFT STANDARD; PRT; 292 AA.
AC Q87A70;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elongation factor Ts (EF-Ts).
TSF OR PD1959.
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22421331; PubMed=12533478;
RX Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furian L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.N., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carter H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Camavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zatz L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa."
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the EF-Ts family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL; AE012560; AAO29789.1; -
 DR HANAP; MF_00050; -; 1.
 DR InterPro; IPR001816; EF_TS.
 DR InterPro; IPR000449; UBA_domain.
 DR Pfam; PF00889; EF_TS; 1.
 DR Pfam; PF00627; UBA; 1.
 DR PROSITE; PS01126; EF_TS 1; 1.
 DR PROSITE; PS01127; EF_TS 2; 1.
 DR Elongation factor; Protein biosynthesis; Complete proteome.
 KW SITE 79 INVOLVED IN MG++ ION DISLOCATION FROM EF-
 FT TU (BY SIMILARITY).
 FT SITE 82

SQ SEQUENCE 292 AA; 31347 MW; 7C2BBA34D56B6E0F CRC64;
 Query Match 2.5%; Score 6; DB 1; Length 292;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIIS 234
 Db 221 LEKIIIS 226

RESULT 146
 VGLG HRSV5 STANDARD; PRT; 298 AA.
 AC P27024;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 35, Last annotation update)
 DE Major surface glycoprotein G (Attachment glycoprotein G).
 GN G.
 OS Human respiratory syncytial virus (strain rsb6190).
 OS Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
 OX NCBI_TaxID=11255;
 RN [1]
 RP SEQUENCE FROM N.A. (PubMed=1895054;
 RX MEDLINE=91374005; PubMed=1895054;
 RA Cane P.A., Matthews D.A., Pringle C.R.;
 RT "Identification of variable domains of the attachment (G) protein of
 RT subgroup A respiratory syncytial viruses.";
 RL J. Gen. Virol. 73:2091-2096(1991).
 CC -!- FUNCTION: Unlike the other paramyxovirus attachment proteins, the
 CC respiratory syncytial virus G protein lacks both neuraminidase and
 CC hemagglutinating activities.
 CC -!- SUBCELLULAR LOCATION: Expressed on the surface of the infected
 CC cells and incorporated in the membrane of the virions.
 CC -!- PTM: May carry 40-80 separate O-linked carbohydrate chains
 CC distributed among the 91 serine and threonine residues.
 DR PIR; JQ1207; JQ1207.
 DR InterPro; IPR000925; Glycoprot G.
 DR Pfam; PF00802; Glycoprotein G; 1.
 KW Transmembrane; Glycoprotein.
 FT DOMAIN 1 37
 FT TRANSMEM 38 66
 FT POTENTIAL.
 FT DOMAIN 67 298
 FT EXTRACELLULAR (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 103 135
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 237 237
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 294 294
 SQ SEQUENCE 298 AA; 32769 MW; 4D74E854D34D7BA5 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 298;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SQTTAI 31
 Db 109 SQTTAI 114
 RESULT 147
 GFR4 HUMAN STANDARD; PRT; 299 AA.
 ID GFR4_HUMAN
 AC Q9GZ27; Q9HI91; Q9HI92;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE GDNF family receptor alpha 4 precursor (GFR-alpha 4) (GFRalpha4)
 DE (persephin receptor).
 GN GFR4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS GFRALPHA4A; GFRALPHA4B AND GFRALPHA4C),
 AND GPI-ANCHOR.
 RC TISSUE=Thyroid;
 RX MEDLINE=21153758; PubMed=11116144;
 RA Lindahl M., Poteryaev D., Yu L., Arumae U., Timmusk T., Bongarzoni I.,
 RA Aiello A., Pierotti M.A., Airaksinen M.S., Saarma M.;
 RT "Human glial cell line-derived neurotrophic factor receptor alpha4 is
 RT the receptor for persephin and is predominantly expressed in normal
 RT and malignant thyroid medullary cells.";
 RL J. Biol. Chem. 276:9344-9351(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM GFRALPHA4A).
 RA Zhou B., Levinson B., Gitschier J.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A. (PubMed=11780052;
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Eutier A.P., Garder C., Carter N.P., Clee C.M.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hamond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharvaisho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
 RA Rice C.M., Ross M.T., Scott C.B., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 CC -!- FUNCTION: Receptor for persephin. Mediates the GDNF-induced
 CC autophosphorylation and activation of the RET receptor. May be
 CC important in C-cell development and, in the postnatal development
 CC of the adrenal medulla.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (isoforms GFRalpha4a and GFRalpha4b). Secreted (isoform
 CC GFRalpha4c).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;


```
CC Comment-Additional isoforms seem to exist;
CC Name=GFRalpha4b;
CC IsoId=Q9GZ27-1; Sequence=Displayed;
CC Name=GFRalpha4a;
CC IsoId=Q9GZ27-2; Sequence=VSP_007223;
CC Name=GFRalpha4c;
CC IsoId=Q9GZ27-3; Sequence=VSP_007224; VSP_007225;
CC TISSUE SPECIFICITY: Predominantly expressed in the adult thyroid
CC gland. Low levels also found in fetal adrenal and thyroid glands.
CC -! SIMILARITY: Belongs to the GDNFR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ291673; CAC19690.1; -
CC EMBL; AJ291674; CAC18691.1; -
CC EMBL; AJ291675; CAC19692.1; -
CC EMBL; AF253318; AAG25925.1; -
CC EMBL; AL356755; CAC16508.2; -
CC GenSeq; HGNC:13821; GFR4.
CC InterPro; IPR003438; GDNF_receptor.
CC Pfam; PF02351; GDNF; 1.
CC PRINTS; PR01316; GDNFRECEPTOR.
CC Receptor; Glycoprotein; GPI-anchor; Membrane; Signal;
KW Alternative splicing; Lipoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 278 GDNF FAMILY RECEPTOR ALPHA 4.
FT PROPEP 279 299 REMOVED IN MATURE FORM (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC...) (POTENTIAL).
FT LIPID 278 278 GPI-anchor amidated glycine (POTENTIAL).
FT VARSPLIC 132 137 CARAAGPWRGNGLSPAHPRAQAQSPGLSLVHPSSQ
FT FT RPRRLPAGPG -> PRLAFOVSCTPAPSPADGCLLQOGR
FT FT APSAPDGLDQAGCLURAYAGLV (in isoform
FT FT GFRalpha4a).
FT FT /FTID=VSP_007223.
FT FT CARAAGPWRGNGRLSPAHPPRAQAQSPGLSLVHPSSQ
FT FT RPRRLPAGPG -> PRLAFOVSCTPAPSPADGCLLQOGR
FT FT CLRAVAGLVSPQAPPSPPLTTWT (in isoform
FT FT GFRalpha4c).
FT FT /FTID=VSP_007224.
FT FT Missing (in isoform GFRalpha4c).
FT FT /FTID=VSP_007225.
FT FT SEQUENCE 299 AA; 31669 MW; 8443B832FF10801 CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 299;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 TGRALE 116
Db 277 TGRALE 282
RESULT 148
CAFR_YERPE
ID CAFR_YERPE STANDARD; PRT; 301 AA.
AC P26950; Q9376;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE F1 operon positive regulatory protein.
GN CAFIR OR YPW1.81C OR Y5193 OR Y1097.
OS Yersinia pestis.
OC Plasmid pMT1 (pMT-1), and Plasmid pFta.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC PLASMID-pFta;
RX MEDLINE=91323540; PubMed=1677900;
RA Galyov E.E., Karlyshev A.V., Chernovskaya T.V., Dolgikh D.A.,
RA Galyov E.E., Karlyshev A.V., Abramov V.M., Zav'yalov V.P.;
RA Smirnov O.Y., Volkovoy K.I., Abramov V.M., Zav'yalov V.P.;
RT "Expression of the envelope antigen F1 of Yersinia pestis is mediated
RT by the product of cafM gene having homology with the chaperrone
RT protein PapD of Escherichia coli.";
RL FEBS Lett. 286:79-82(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID-pFta;
RX MEDLINE=92339520; PubMed=1633857;
RA Karlyshev A.V., Galyov E.E., Abramov V.M., Zav'yalov V.P.;
RT "CafR gene and its role in the regulation of capsule formation of Y.
RT pestis.";
RL FEBS Lett. 305:37-40(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=99043898; PubMed=9826348;
RA Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT KIM5 plasmid encoding murine toxin and capsular antigen.";
RL Infect. Immun. 66:5731-5742(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCreedy P., Skowronski E., Garnes J.,
RA Kobayashi A., Brubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Yersinia
RT pestis.";
RL J. Bacteriol. 180:5192-5202(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-32 / Biovar Orientalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -! FUNCTION: POSITIVE REGULATOR OF F1 OPERON EXPRESSION.
CC -! SIMILARITY: BELONGS TO THE ARAC/YXIS FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61996; CAA43969.1; -
CC EMBL; AF074611; AAC82755.1; ALT_INIT.
CC EMBL; AF053947; AAC13221.1; -
CC EMBL; AL117211; CAB55263.1; ALT_INIT.
CC PIR; S19097; S19097.
CC PIR; T14705; T14705.
CC HSP; P27246; IBL0.
CC InterPro; IPR000005; HTHARAC.
CC Pfam; PF00165; HTH_Arac; 2.
CC PRINTS; PR00032; HTHARAC.
CC SMART; SM00342; HTH_ARAC; 1.
CC PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
CC PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
CC Transcription regulation; Activator; DNA-binding; Plasmid;
KW
```

```
KW Complete proteome.
FT DNA_BIND 25 44 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 124 124 E -> V (IN REF. 1).
FT CONFLICT 135 135 E -> V (IN REF. 1).
SQ SEQUENCE 301 AA; 36053 MW; E1EB5AD5C4CD43C0 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 LYSGFS 46
DB 30 LYSGFS 35

RESULT 149
Y548 STAPB STANDARD; PRT; 302 AA.
AC Q5CTE3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0042 protein SE0548.
GN SE0548.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qian Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -1- SIMILARITY: Belongs to the UPF0042 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016745; AAO04145.1; -.
CC HAMAP; MF_00636; -.
CC InterPro; IPR005337; UPF0042.
CC Pfam; PF03668; ATP_bind2; 1.
CC Hypothetical protein; ATP-binding; Complete proteome.
FT NP_BIND 18 25 ATP (POTENTIAL)
SQ SEQUENCE 302 AA; 34751 MW; 4516DEBA894DA5F CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EKISR 235
DB 104 EKISR 109

RESULT 150
Y367 RICPR STANDARD; PRT; 303 AA.
AC Q9ZDG2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP367.
GN RP367.
```

```
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sierhitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria";
RL Nature 396:133-140(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ235271; CAA14826.1; -.
CC PIR; H71693; H71693.
CC InterPro; IPR007487; DUF534.
CC Pfam; PF04392; DUF534; 1.
CC Hypothetical protein; Complete proteome.
SQ SEQUENCE 303 AA; 33569 MW; D7703DEB5F29440 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 PSFLFV 50
DB 11 PSFLFV 16

Search completed: April 5, 2004, 07:37:55
Job time : 24 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: April 5, 2004, 07:30:34 ; Search time 45 Seconds
(without alignments)
1710.810 Million cell updates/sec

Title: US-10-032-221B-10
Perfect score: 244
Sequence: 1 GLKRGKDSGSPATWTRGF.....KAGELEKIIISRCQVCMKKRH 244

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rhodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 244 | 100.0 | 245 | 4 Q9NYC4 | Q9NYC4 homo sapien |
| 2 | 117 | 48.0 | 212 | 6 Q28512 | Q28512 macaca mula |
| 3 | 54 | 22.1 | 203 | 6 Q28682 | Q28682 oryctolagus |
| 4 | 39 | 16.0 | 112 | 11 Q8CCD6 | Q8CCD6 mus musculus |
| 5 | 39 | 16.0 | 230 | 11 Q61122 | Q61122 rattus norv |
| 6 | 39 | 16.0 | 246 | 11 Q61435 | Q61435 mus musculus |
| 7 | 39 | 16.0 | 1669 | 11 Q9QZS0 | Q9QZS0 mus musculus |
| 8 | 38 | 15.6 | 161 | 11 Q61430 | Q61430 mus musculus |
| 9 | 36 | 14.8 | 52 | 4 Q9UDK9 | Q9UDK9 homo sapien |
| 10 | 36 | 14.8 | 56 | 4 Q9UDL0 | Q9UDL0 homo sapien |
| 11 | 36 | 14.8 | 97 | 4 Q9UDL1 | Q9UDL1 homo sapien |
| 12 | 35 | 14.3 | 212 | 6 Q28567 | Q28567 ovis aries |
| 13 | 25 | 10.2 | 210 | 6 Q28273 | Q28273 canis famli |
| 14 | 24 | 9.8 | 203 | 6 Q29032 | Q29032 sus scrofa |
| 15 | 18 | 7.4 | 29 | 4 Q9UDJ9 | Q9UDJ9 homo sapien |
| 16 | 17 | 7.0 | 179 | 11 P70165 | P70165 mus musculus |

| | | | | | |
|----|-----|------|----|--------|---------------------|
| 17 | 7.0 | 225 | 6 | Q28271 | Q28271 canis famli |
| 18 | 7.0 | 226 | 11 | Q99LQ8 | Q99LQ8 mus musculus |
| 19 | 7.0 | 229 | 4 | Q9NF88 | Q9NF88 homo sapien |
| 20 | 7.0 | 229 | 4 | Q9NYC5 | Q9NYC5 homo sapien |
| 21 | 7.0 | 253 | 11 | Q61436 | Q61436 mus musculus |
| 22 | 7.0 | 585 | 11 | Q8OV57 | Q8OV57 mus musculus |
| 23 | 7.0 | 799 | 11 | Q8BNS7 | Q8BNS7 mus musculus |
| 24 | 7.0 | 886 | 4 | Q9NUB7 | Q9NUB7 homo sapien |
| 25 | 7.0 | 979 | 13 | Q919K3 | Q919K3 gallus gall |
| 26 | 7.0 | 1075 | 4 | Q8EX41 | Q8EX41 homo sapien |
| 27 | 7.0 | 1621 | 4 | Q9H4R9 | Q9H4R9 homo sapien |
| 28 | 7.0 | 1684 | 6 | Q8HYC1 | Q8HYC1 canis famli |
| 29 | 7.0 | 1688 | 6 | Q866Z2 | Q866Z2 canis famli |
| 30 | 7.0 | 1691 | 11 | Q9ESQ2 | Q9ESQ2 mus musculus |
| 31 | 4.9 | 1752 | 5 | Q87265 | Q87265 strongyloce |
| 32 | 4.5 | 1713 | 5 | Q9GV24 | Q9GV24 sarcophaga |
| 33 | 4.5 | 1761 | 5 | Q18407 | Q18407 drosophila |
| 34 | 4.5 | 1940 | 5 | Q9VMV5 | Q9VMV5 drosophila |
| 35 | 10 | 673 | 4 | Q14052 | Q14052 homo sapien |
| 36 | 10 | 1024 | 5 | Q8T7S4 | Q8T7S4 anopheles g |
| 37 | 4.1 | 1747 | 5 | Q26640 | Q26640 strongyloce |
| 38 | 3.7 | 59 | 11 | Q8BP19 | Q8BP19 mus musculus |
| 39 | 3.7 | 312 | 11 | Q84457 | Q84457 mus musculus |
| 40 | 3.7 | 358 | 11 | Q91V13 | Q91V13 mus musculus |
| 41 | 3.7 | 1682 | 11 | Q9QZK9 | Q9QZK9 mus musculus |
| 42 | 8 | 74 | 2 | Q48996 | Q48996 mycoplasma |
| 43 | 3.3 | 179 | 5 | Q8T2W5 | Q8T2W5 trypanosoma |
| 44 | 3.3 | 304 | 16 | Q81T03 | Q81T03 bacillus an |
| 45 | 3.3 | 337 | 11 | Q8BX78 | Q8BX78 mus musculus |
| 46 | 3.3 | 337 | 11 | Q8BGP6 | Q8BGP6 mus musculus |
| 47 | 3.3 | 452 | 11 | Q921S0 | Q921S0 mus musculus |
| 48 | 3.3 | 486 | 11 | Q8BN95 | Q8BN95 mus musculus |
| 49 | 3.3 | 523 | 17 | Q8TH26 | Q8TH26 methanosarc |
| 50 | 3.3 | 541 | 11 | Q8C7W7 | Q8C7W7 mus musculus |
| 51 | 3.3 | 546 | 11 | Q99K97 | Q99K97 mus musculus |
| 52 | 3.3 | 1477 | 3 | Q8C250 | Q8C250 neurospora |
| 53 | 3.3 | 1691 | 11 | Q9ESQ1 | Q9ESQ1 mus musculus |
| 54 | 3.3 | 1723 | 5 | Q9QCB1 | Q9QCB1 hydra atten |
| 55 | 3.3 | 1779 | 5 | Q9VMV4 | Q9VMV4 drosophila |
| 56 | 2.9 | 49 | 2 | Q9R4U1 | Q9R4U1 azotobacter |
| 57 | 2.9 | 68 | 16 | Q8XW59 | Q8XW59 ralstonia s |
| 58 | 2.9 | 69 | 2 | Q84ER2 | Q84ER2 ralstonia o |
| 59 | 2.9 | 109 | 12 | Q67050 | Q67050 influenzavi |
| 60 | 2.9 | 109 | 12 | Q67053 | Q67053 influenzavi |
| 61 | 2.9 | 109 | 12 | Q67051 | Q67051 influenzavi |
| 62 | 2.9 | 109 | 12 | Q67052 | Q67052 influenzavi |
| 63 | 2.9 | 116 | 16 | Q89PD4 | Q89PD4 bradyrhizob |
| 64 | 2.9 | 118 | 16 | Q8EKD6 | Q8EKD6 shewanella |
| 65 | 2.9 | 122 | 17 | Q9YEN1 | Q9YEN1 aeropyrum p |
| 66 | 2.9 | 125 | 2 | Q68196 | Q68196 klebsiella |
| 67 | 2.9 | 128 | 17 | Q96Y71 | Q96Y71 sulfolobus |
| 68 | 2.9 | 157 | 8 | Q9MJR3 | Q9MJR3 taenia pisi |
| 69 | 2.9 | 168 | 17 | Q9YDQ8 | Q9YDQ8 aeropyrum p |
| 70 | 2.9 | 171 | 5 | Q97232 | Q97232 plasmodium |
| 71 | 2.9 | 173 | 8 | Q8HKQ0 | Q8HKQ0 apasama min |
| 72 | 2.9 | 174 | 16 | Q82WJ0 | Q82WJ0 nitrosomonas |
| 73 | 2.9 | 175 | 5 | Q8SVV5 | Q8SVV5 encephalito |
| 74 | 2.9 | 182 | 4 | Q8ND70 | Q8ND70 homo sapien |
| 75 | 2.9 | 187 | 17 | Q97ZW3 | Q97ZW3 sulfolobus |
| 76 | 2.9 | 189 | 2 | Q59673 | Q59673 propionibac |
| 77 | 2.9 | 193 | 2 | Q9ZGN1 | Q9ZGN1 azotobacter |
| 78 | 2.9 | 199 | 12 | Q8TR59 | Q8TR59 human reapi |
| 79 | 2.9 | 200 | 16 | Q9A303 | Q9A303 caulobacter |
| 80 | 2.9 | 202 | 6 | Q28272 | Q28272 canis famli |
| 81 | 2.9 | 202 | 16 | Q83HP6 | Q83HP6 tropheryma |
| 82 | 2.9 | 202 | 16 | Q83GI4 | Q83GI4 tropheryma |
| 83 | 2.9 | 205 | 6 | Q28274 | Q28274 canis famli |
| 84 | 2.9 | 206 | 3 | Q96UT6 | Q96UT6 candida alb |
| 85 | 2.9 | 208 | 6 | Q29468 | Q29468 canis famli |
| 86 | 2.9 | 211 | 17 | Q96184 | Q96184 sulfolobus |
| 87 | 2.9 | 216 | 2 | P95356 | P95356 neisseria g |
| 88 | 2.9 | 216 | 16 | Q9K1R3 | Q9K1R3 neisseria m |
| 89 | 2.9 | 216 | 16 | Q9JW71 | Q9JW71 neisseria m |

| | | | | | | | | | | | | | |
|-----|---|-----|-----|----|--------|---------------------|-----|---|-----|-----|----|--------|--------------------|
| 90 | 7 | 2.9 | 220 | 5 | Q8MU16 | Q8mu16 trichinella | 163 | 7 | 2.9 | 471 | 12 | Q9E311 | Q9e311 influenza a |
| 91 | 7 | 2.9 | 233 | 4 | Q8IXG7 | Q8ixg7 homo sapien | 164 | 7 | 2.9 | 471 | 12 | Q9E313 | Q9e313 influenza a |
| 92 | 7 | 2.9 | 236 | 16 | Q7V4W7 | Q7v4w7 prochloroco | 165 | 7 | 2.9 | 472 | 2 | Q8GBR8 | Q8gbr8 treponema m |
| 93 | 7 | 2.9 | 237 | 16 | Q9K8B8 | Q9k8b8 bacillus ha | 166 | 7 | 2.9 | 472 | 12 | Q9E310 | Q9e310 influenza a |
| 94 | 7 | 2.9 | 239 | 16 | Q8F8X2 | Q8f8x2 leptospira | 167 | 7 | 2.9 | 473 | 12 | Q9E315 | Q9e315 influenza a |
| 95 | 7 | 2.9 | 241 | 5 | Q8ML71 | Q8ml71 drosophila | 168 | 7 | 2.9 | 477 | 4 | Q9BUD0 | Q9bud0 homo sapien |
| 96 | 7 | 2.9 | 244 | 4 | Q8N2M3 | Q8n2m3 homo sapien | 169 | 7 | 2.9 | 478 | 16 | Q9WZ22 | Q9wz22 thermotoga |
| 97 | 7 | 2.9 | 245 | 16 | Q8G7Z6 | Q8g7z6 bifidobacte | 170 | 7 | 2.9 | 494 | 4 | Q8H0M4 | Q8h0m4 homo sapien |
| 98 | 7 | 2.9 | 247 | 16 | Q8XEX8 | Q8xex8 salmonella | 171 | 7 | 2.9 | 503 | 12 | Q8JL66 | Q8jl66 ectromelia |
| 99 | 7 | 2.9 | 252 | 16 | Q87BW1 | Q87bw1 xyella fas | 172 | 7 | 2.9 | 503 | 12 | Q7Z739 | Q7z739 cowpox viru |
| 100 | 7 | 2.9 | 254 | 16 | Q81YK7 | Q81yk7 bacillus an | 173 | 7 | 2.9 | 511 | 10 | Q8LRA0 | Q8lra0 cryza sativ |
| 101 | 7 | 2.9 | 254 | 16 | Q81AT7 | Q81at7 bacillus ce | 174 | 7 | 2.9 | 523 | 4 | Q99993 | Q99993 homo sapien |
| 102 | 7 | 2.9 | 255 | 10 | Q9ZWM8 | Q9zwm8 cryza sativ | 175 | 7 | 2.9 | 548 | 12 | Q919W4 | Q919w4 influenza a |
| 103 | 7 | 2.9 | 257 | 16 | Q9PB46 | Q9pb46 xyella fas | 176 | 7 | 2.9 | 548 | 12 | Q919W5 | Q919w5 influenza a |
| 104 | 7 | 2.9 | 259 | 8 | Q8SM64 | Q8sm64 toxoplasma | 177 | 7 | 2.9 | 548 | 12 | Q919W3 | Q919w3 influenza a |
| 105 | 7 | 2.9 | 259 | 10 | Q9SNQ0 | Q9snq0 cryza sativ | 178 | 7 | 2.9 | 549 | 6 | Q86423 | Q86423 bos taurus |
| 106 | 7 | 2.9 | 264 | 16 | Q34535 | Q34535 bacillus su | 179 | 7 | 2.9 | 550 | 12 | Q8QLT9 | Q8qlt9 influenza a |
| 107 | 7 | 2.9 | 267 | 16 | Q92C22 | Q92c22 listeria in | 180 | 7 | 2.9 | 550 | 12 | Q98VJ6 | Q98vu6 influenza a |
| 108 | 7 | 2.9 | 267 | 16 | Q8Y880 | Q8y880 listeria mo | 181 | 7 | 2.9 | 550 | 12 | Q82499 | Q82499 influenzavi |
| 109 | 7 | 2.9 | 267 | 16 | Q8ET45 | Q8et45 oceanobacil | 182 | 7 | 2.9 | 550 | 12 | Q82498 | Q82498 influenzavi |
| 110 | 7 | 2.9 | 274 | 16 | Q7V2M5 | Q7v2m5 prochloroco | 183 | 7 | 2.9 | 550 | 12 | Q82753 | Q82753 influenzavi |
| 111 | 7 | 2.9 | 282 | 10 | Q9LWS3 | Q9lws3 cryza sativ | 184 | 7 | 2.9 | 565 | 12 | Q67103 | Q67103 influenza a |
| 112 | 7 | 2.9 | 285 | 17 | Q8Q083 | Q8q083 methanosarc | 185 | 7 | 2.9 | 565 | 12 | Q67107 | Q67107 influenza a |
| 113 | 7 | 2.9 | 297 | 16 | Q8F0K1 | Q8f0k1 leptospira | 186 | 7 | 2.9 | 565 | 12 | Q67104 | Q67104 influenza a |
| 114 | 7 | 2.9 | 304 | 16 | Q88GA3 | Q88ga3 pseudomonas | 187 | 7 | 2.9 | 565 | 12 | Q67102 | Q67102 influenza a |
| 115 | 7 | 2.9 | 313 | 11 | Q8C7L9 | Q8c7l9 mus muscullu | 188 | 7 | 2.9 | 565 | 12 | Q91B65 | Q91b65 influenza v |
| 116 | 7 | 2.9 | 313 | 17 | Q8Z2Y2 | Q8z2y2 pyrobaculum | 189 | 7 | 2.9 | 565 | 12 | Q91B65 | Q91b65 influenza v |
| 117 | 7 | 2.9 | 316 | 16 | Q8JL58 | Q8jl58 rhizobium l | 190 | 7 | 2.9 | 565 | 12 | Q82559 | Q82559 influenza a |
| 118 | 7 | 2.9 | 331 | 16 | Q88B27 | Q88b27 pseudomonas | 191 | 7 | 2.9 | 565 | 12 | Q82793 | Q82793 influenzavi |
| 119 | 7 | 2.9 | 338 | 16 | Q92T66 | Q92t66 rhizobium m | 192 | 7 | 2.9 | 565 | 12 | Q82793 | Q82793 influenzavi |
| 120 | 7 | 2.9 | 343 | 10 | Q9SAC7 | Q9sac7 arabidopsis | 193 | 7 | 2.9 | 565 | 12 | Q91E71 | Q91e71 influenzavi |
| 121 | 7 | 2.9 | 354 | 16 | Q9ISY7 | Q9isy7 pseudomonas | 194 | 7 | 2.9 | 565 | 12 | Q82792 | Q82792 influenzavi |
| 122 | 7 | 2.9 | 365 | 12 | Q56962 | Q56962 influenza a | 195 | 7 | 2.9 | 565 | 12 | Q67106 | Q67106 influenza a |
| 123 | 7 | 2.9 | 365 | 12 | Q56961 | Q56961 influenza a | 196 | 7 | 2.9 | 565 | 12 | Q66752 | Q66752 equine infl |
| 124 | 7 | 2.9 | 371 | 12 | Q9DL24 | Q9dl24 influenza a | 197 | 7 | 2.9 | 565 | 12 | Q86639 | Q86639 equine infl |
| 125 | 7 | 2.9 | 372 | 4 | Q9NT4 | Q9nt4 homo sapien | 198 | 7 | 2.9 | 566 | 12 | Q98052 | Q98052 influenzavi |
| 126 | 7 | 2.9 | 372 | 4 | Q9N3J5 | Q9n3j5 homo sapien | 199 | 7 | 2.9 | 566 | 12 | Q37160 | Q37160 influenzavi |
| 127 | 7 | 2.9 | 372 | 4 | Q8IUZ7 | Q8iuz7 homo sapien | 200 | 7 | 2.9 | 566 | 12 | Q82496 | Q82496 influenzavi |
| 128 | 7 | 2.9 | 382 | 10 | Q8H666 | Q8h666 cryza sativ | | | | | | | |
| 129 | 7 | 2.9 | 384 | 12 | Q8JX63 | Q8jx63 influenza a | | | | | | | |
| 130 | 7 | 2.9 | 389 | 10 | Q7XSV3 | Q7xsv3 cryza sativ | | | | | | | |
| 131 | 7 | 2.9 | 398 | 16 | Q89C31 | Q89c31 bradyrhizob | | | | | | | |
| 132 | 7 | 2.9 | 403 | 4 | Q9NWF7 | Q9nwf7 homo sapien | | | | | | | |
| 133 | 7 | 2.9 | 407 | 16 | Q8EG37 | Q8eg37 shewanella | | | | | | | |
| 134 | 7 | 2.9 | 409 | 12 | Q9Q0L5 | Q9q0l5 influenza a | | | | | | | |
| 135 | 7 | 2.9 | 412 | 2 | Q93S15 | Q93is burkholderi | | | | | | | |
| 136 | 7 | 2.9 | 412 | 2 | Q89122 | Q89l22 burkholderi | | | | | | | |
| 137 | 7 | 2.9 | 416 | 2 | Q8KN89 | Q8kn89 pseudomonas | | | | | | | |
| 138 | 7 | 2.9 | 416 | 12 | Q8KN80 | Q8kn80 pseudomonas | | | | | | | |
| 139 | 7 | 2.9 | 419 | 13 | Q92043 | Q92043 crotales at | | | | | | | |
| 140 | 7 | 2.9 | 424 | 5 | Q9GSD3 | Q9gsd3 plasmodium | | | | | | | |
| 141 | 7 | 2.9 | 424 | 5 | Q8T9R7 | Q8t9r7 plasmodium | | | | | | | |
| 142 | 7 | 2.9 | 424 | 5 | Q9GSD7 | Q9gsd7 plasmodium | | | | | | | |
| 143 | 7 | 2.9 | 424 | 5 | Q9NH61 | Q9nh61 plasmodium | | | | | | | |
| 144 | 7 | 2.9 | 424 | 5 | Q9NH61 | Q9nh61 plasmodium | | | | | | | |
| 145 | 7 | 2.9 | 424 | 5 | Q9NH61 | Q9nh61 plasmodium | | | | | | | |
| 146 | 7 | 2.9 | 424 | 5 | Q8GM70 | Q8gm70 plasmodium | | | | | | | |
| 147 | 7 | 2.9 | 424 | 5 | Q8GM70 | Q8gm70 plasmodium | | | | | | | |
| 148 | 7 | 2.9 | 424 | 5 | Q8GM70 | Q8gm70 plasmodium | | | | | | | |
| 149 | 7 | 2.9 | 425 | 10 | Q9ST79 | Q9st79 cryza sativ | | | | | | | |
| 150 | 7 | 2.9 | 429 | 12 | Q9Q0L4 | Q9q0l4 influenza a | | | | | | | |
| 151 | 7 | 2.9 | 438 | 12 | Q9Q0L3 | Q9q0l3 influenza a | | | | | | | |
| 152 | 7 | 2.9 | 438 | 16 | Q99V88 | Q99vd8 staphylococ | | | | | | | |
| 153 | 7 | 2.9 | 438 | 16 | Q8CT88 | Q8ct88 staphylococ | | | | | | | |
| 154 | 7 | 2.9 | 442 | 10 | Q9SGX5 | Q9sgx5 arabidopsis | | | | | | | |
| 155 | 7 | 2.9 | 449 | 17 | Q9PT71 | Q9pt71 methanosarc | | | | | | | |
| 156 | 7 | 2.9 | 450 | 16 | Q7TW4 | Q7tw4 mycobacteri | | | | | | | |
| 157 | 7 | 2.9 | 451 | 12 | Q98173 | Q98l73 molluscum c | | | | | | | |
| 158 | 7 | 2.9 | 451 | 12 | Q98173 | Q98l73 molluscum c | | | | | | | |
| 159 | 7 | 2.9 | 459 | 16 | Q8X151 | Q8x151 clostridium | | | | | | | |
| 160 | 7 | 2.9 | 466 | 11 | Q80WCS | Q80wcs mus muscullu | | | | | | | |
| 161 | 7 | 2.9 | 467 | 10 | Q84JS7 | Q84js7 arabidopsis | | | | | | | |
| 162 | 7 | 2.9 | 467 | 12 | Q9E312 | Q9e312 influenza a | | | | | | | |
| | | | 470 | 12 | Q9E314 | Q9e314 influenza a | | | | | | | |

ALIGNMENTS

RESULT 1

| | |
|--------|--|
| Q9NYC4 | PRELIMINARY; PRT; 245 AA. |
| ID | Q9NYC4 |
| AC | Q9NYC4; |
| DT | 01-OCT-2000 (trEMBLrel. 15, Created) |
| DT | 01-OCT-2000 (trEMBLrel. 15, Last sequence update) |
| DT | 01-OCT-2003 (trEMBLrel. 25, Last annotation update) |
| DE | Tumstatin (Fragment). |
| OS | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| OX | NCBI_TaxID:9606; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RA | Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A., |
| RA | Ericksen M.D., Hoffer H., Xiao Y., Stillman I.E., Kalluri R.; |
| RT | "Distinct anti-tumor properties of a type IV collagen domain derived |
| RT | from basement membrane."; |
| RL | J. Biol. Chem. 0:0-0(2000). |
| DR | EMBL; AF259351; AAF7632.1; |
| DR | GO; GO:0005581; C:collagen; IEA. |
| DR | GO; GO:0005201; F:extracellular matrix structural constituent; IEA. |
| DR | GO; GO:0003676; F:nucleic acid binding; IEA. |
| DR | InterPro; IPR001442; Procollag4 C. |
| DR | InterPro; IPR000504; RNA_rec_mot. |
| DR | Pfam; PF01413; C4; 2. |
| DR | ProDom; PD003923; ProcollagNC4; 1. |
| DR | SMART; SM00111; C4; 2. |
| DR | PROSITE; PS00030; RRM_RNP_1; 1. |
| FT | NON_TER 1 |

SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;
Query Match 100.0%; Score 244; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.6e-242; Indels 0; Gaps 0;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGPATWTRGFTVTHSQTTPSCPTVPLYSQFSLFVQGNQRAHQD 60
Db 2 GLKGRGDSGPATWTRGFTVTHSQTTPSCPTVPLYSQFSLFVQGNQRAHQD 61

Qy 61 LGTSGSLQRTTWPFLFCNVNDYCNFASRNDYSYWLSTPALPMNMAPITGRALPEYIS 120
Db 62 LGTSGSLQRTTWPFLFCNVNDYCNFASRNDYSYWLSTPALPMNMAPITGRALPEYIS 121

Qy 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPIMFTSAGSEGTQALASPGSCLE 180
Db 122 RCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPIMFTSAGSEGTQALASPGSCLE 181

Qy 181 EFRASPFLECHGRGTCNYNSYSFWSLASLNPFRKPIPTSVKAGELEKIISRCQVCM 240
Db 182 EFRASPFLECHGRGTCNYNSYSFWSLASLNPFRKPIPTSVKAGELEKIISRCQVCM 241

Qy 241 KKRH 244
Db 242 KKRH 245

RESULT 2
Q28512 PRELIMINARY; PRT; 212 AA.
AC Q28512;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; L47280; AAA91861.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;

Query Match 48.0%; Score 117; DB 6; Length 212;
Best Local Similarity 100.0%; Pred. No. 7.1e-112;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAVHSQTTDIPPCPHGWSLWKGFSPIMFTSAGSEGTQALASPGSCLEFPASGPF 187
Db 96 PAIAIAVHSQTTDIPPCPHGWSLWKGFSPIMFTSAGSEGTQALASPGSCLEFPASGPF 155

Qy 188 LECHGRGTCNYNSYSFWSLASLNPFRKPIPTSVKAGELEKIISRCQVCMKKRH 244

Db 156 LECHGRGTCNYNSYSFWSLASLNPFRKPIPTSVKAGELEKIISRCQVCMKKRH 212

RESULT 3
Q28682 PRELIMINARY; PRT; 203 AA.
AC Q28682;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Cricetidae;
OX NCBI_TaxID=9986;
RN [1]
RP TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; L47283; AAA91893.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 22.1%; Score 54; DB 6; Length 203;
Best Local Similarity 100.0%; Pred. No. 4.7e-47;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPIMFTSAGSEG 167
Db 82 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPIMFTSAGSEG 135

RESULT 4
Q8CCD6 PRELIMINARY; PRT; 112 AA.
AC Q8CCD6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Procollagen (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK033387; BAC28260.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

```

DR GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
SQ SEQUENCE 112 AA; 12438 MW; FFOFDD3C95ATBF31 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 112;
Best Local Similarity 100.0%; Pred. No. 7.8e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAG 227
DB 57 ECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAG 95

RESULT 5
O63122
ID Q63122 PRELIMINARY; PRT; 230 AA.
AC Q63122;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RX MEDLINE=98210005; PubMed=9550634;
RA Ryan J.J., Katbama I., Mason P.J., Pusey C.D., Turner A.N.;
RT "Sequence analysis of the 'Goopasture antigen' of mammals.";
RL Nephrol. Dial. Transplant. 13:602-607(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA Turner N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U47281; AAB72238.2; -.
DR GO:0005581; F:extracellular matrix structural constituent; IEA.
DR GO:0005201; F:nucleic acid binding; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.4e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAG 227
DB 175 ECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAG 213

RESULT 6
O61435
ID Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA Turner N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U47281; AAB72238.2; -.
DR GO:0005581; F:extracellular matrix structural constituent; IEA.
DR GO:0005201; F:nucleic acid binding; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.4e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAG 227
DB 175 ECHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAG 213

RESULT 7
O9QZS0
ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.
AC Q9QZS0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 3 collagen IV.
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169387; AAD50449.1; -.
DR PIR; I48302; I48302.
DR MGD; MGI:104688; Col4a3.
DR GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.

```

```

DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 21.
DR Prodom: PD000007; Cig_helix; 6.
DR Prodom: PD003923; ProcollagenC4; 1.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 1669;
Best Local Similarity 100.0%; Pred. No. 7.9e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFKPIPSIVKAG 227
DB 1614 ECHGRGTCNYNSYSFWLASLNPERMFKPIPSIVKAG 1652

RESULT 8
Q61430
ID O61430 PRELIMINARY; PRT; 161 AA.
AC O61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbaumer I.;
RT "Cloning of the Ncl domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells."
RL Submitted (OCT-1994) to the EMBL/GenBank/DDER databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0008201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4_C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
DR NON_TER 161
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFB8236C5 CRC64;

Query Match 15.6%; Score 38; DB 11; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFKPIPSIVKAG 226
DB 124 ECHGRGTCNYNSYSFWLASLNPERMFKPIPSIVKAG 161

RESULT 9
Q9UDK9
ID Q9UDK9 PRELIMINARY; PRT; 52 AA.
AC Q9UDK9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Penades J.R., Bernal D., Revert F., Johansson C., Fresquet V.J.,
RA Cervera J., Wieslander J., Quinones S., Saus J.;
RT "Characterization and expression of multiple alternatively spliced
RT transcripts of the Goodpasture antigen gene region. Goodpasture
RT antibodies recognize recombinant proteins representing the autoantigen
RT and one of its alternative forms."
RL Eur. J. Biochem. 229:754-760(1995).
DR PIR; S69113; S69113.
DR NON_TER 1
FT NON_TER 1
FT NON_CONS 45
FT NON_TER 52
FT NON_TER 52
SQ SEQUENCE 52 AA; 5442 MW; 046AB41B149DDAE3 CRC64;

Query Match 14.8%; Score 36; DB 4; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.9e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKGDGSGPATWTRGTFVTRHSQTTAIPSCPE 36
DB 10 GLKGKGDGSGPATWTRGTFVTRHSQTTAIPSCPE 45

RESULT 10
Q9UDLO
ID Q9UDLO PRELIMINARY; PRT; 56 AA.
AC Q9UDLO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Penades J.R., Bernal D., Revert F., Johansson C., Fresquet V.J.,
RA Cervera J., Wieslander J., Quinones S., Saus J.;
RT "Characterization and expression of multiple alternatively spliced
RT transcripts of the Goodpasture antigen gene region. Goodpasture
RT antibodies recognize recombinant proteins representing the autoantigen
RT and one of its alternative forms."
RL Eur. J. Biochem. 229:754-760(1995).
DR PIR; A49736; A49736.
DR NON_TER 1
FT NON_TER 1
FT NON_CONS 45
FT NON_TER 56
FT NON_TER 56
SQ SEQUENCE 56 AA; 5813 MW; 6A6A2E2819F473B CRC64;

Query Match 14.8%; Score 36; DB 4; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.2e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKGDGSGPATWTRGTFVTRHSQTTAIPSCPE 36
DB 10 GLKGKGDGSGPATWTRGTFVTRHSQTTAIPSCPE 45

RESULT 11
Q9UDL1
ID Q9UDL1 PRELIMINARY; PRT; 97 AA.
AC Q9UDL1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95278230; PubMed=7758473;
RA Perades J.R., Bernal D., Revert F., Johansson C., Fresquet V.J.,
RA Cervera J., Wieselander J., Quinones S., Saus J.;
RT "Characterization and expression of multiple alternatively spliced
RT transcripts of the Goodpasture antigen gene region. Goodpasture
RT antibodies recognize recombinant proteins representing the autoantigen
RT and one of its alternative forms.";
RL Eur. J. Biochem. 229:754-760(1995).
DR PIR; B49736; B49736.
FT NON_TER 1
FT NON_TER 45
FT NON_TER 97
FT NON_TER 97
SQ SEQUENCE 97 AA; 10559 MW; B3CCL127F24A27E3 CRC64;

Query Match 14.8%; Score 36; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.4e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPE 36
DB 10 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPE 45

RESULT 12
Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 14.3%; Score 35; DB 6; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.8e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 GWISLWKGFSPFMFTSAGSEGTGQALAGPGSCLEE 181
DB 115 GWISLWKGFSPFMFTSAGSEGTGQALAGPGSCLEE 149

RESULT 13
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119B4C823633D CRC64;

Query Match 10.2%; Score 25; DB 6; Length 210;
Best Local Similarity 100.0%; Pred. No. 3.4e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 AHGQDLGTGLGSCLOQRTTTPFLFCN 80
DB 34 AHGQDLGTGLGSCLOQRTTTPFLFCN 58

RESULT 14
Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 23025 MW; 31119B4C823633D CRC64;
```


DR ProDom: PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
KW Collagen.
FT NON_TER 203 1
FT NON_TER 203 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 9.8%; Score 24; DB 6; Length 203;
Best Local Similarity 100.0%; Pred. No. 3.5e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 HGQDLGTGLGSLQRFTHMPFLFCN 80
|||||
DB 25 HGQDLGTGLGSLQRFTHMPFLFCN 48
|||||

RESULT 15

Q9UDJ9 PRELIMINARY; PRT; 29 AA.
AC Q9UDJ9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced".
RL J. Biol. Chem. 268:12090-12094 (1993).
FT NON_TER 18 1
FT NON_TER 18 19
FT NON_TER 29 29
SQ SEQUENCE 29 AA; 3102 MW; 2B7047AAB1580036 CRC64;

Query Match 7.4%; Score 18; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTR 18
|||||
DB 1 GLKGRGDSGSPATWTR 18
|||||

RESULT 16

P70165 PRELIMINARY; PRT; 179 AA.
ID P70165
AC P70165;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Collagen type IV alpha5 chain (Fragment).
OS COL4A5.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Oberbauer I.;
RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82218; CAA57698.1;
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
GO; GO:0005581; C:collagen; IEA.

DR GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER 179 179
FT NON_TER 179 179
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 7.0%; Score 17; DB 11; Length 179;
Best Local Similarity 100.0%; Pred. No. 5.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
|||||
DB 47 VCNFASRNDYSYWLSTP 63
|||||

RESULT 17

Q28271 PRELIMINARY; PRT; 225 AA.
ID Q28271;
AC Q28271;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OC Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Samoyed;
RA MEDLINE=9678820; PubMed=8662866;
RT "Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Samoyed;
RA Thorner P.S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50933; AAC48583.2;
GO; GO:0005581; C:collagen; IEA.
DR GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 225 225
SQ SEQUENCE 225 AA; 24585 MW; 2C20455890416E47 CRC64;

Query Match 7.0%; Score 17; DB 6; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
|||||
DB 76 VCNFASRNDYSYWLSTP 92
|||||

RESULT 18

Q99LQ8 PRELIMINARY; PRT; 226 AA.
ID Q99LQ8;
AC Q99LQ8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

```

DE Hypothetical protein (Fragment).
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002269; AA02269.1; -.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 226 AA; 25042 MW; 4F7F0D5371181C21 CRC64;

Query Match 7.0%; Score 17; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 67 VCNFASRNDYSYWLSTP 83

RESULT 19
Q9NF88
ID Q9NF88 PRELIMINARY; PRT; 229 AA.
AC Q9NF88
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RA He A.B.;
RL "Cloning and Expression of Arresten in Escherichia coli and Pachia
pastoris."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF536207; AA097359.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 229 AA; 25391 MW; 09B21FDSA3517B9E CRC64;

Query Match 7.0%; Score 17; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 70 VCNFASRNDYSYWLSTP 86

RESULT 20
Q9NYC5
ID Q9NYC5 PRELIMINARY; PRT; 229 AA.
AC Q9NYC5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H.,
Takanashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K.,
Brickson M.B., Dhanabal M., Simons M., Post M., Kufe D.,
Weichselbaum R.R., Sukhatme V.P., Kalluri R.;
RT "Anti-angiogenic cues from vascular basement membrane collagen."
RN [2]
RN Cancer Res. 0:0-0(2000).
RP SEQUENCE FROM N.A.
RA Fu J., Bai X., Wang W., Ruan C.;
RT "Arresten, a collagen-derived inhibitor of angiogenesis."
RL Chung Hua Hsueh Yeh Hsueh Tsa Chih 22:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Peng X., Yin B., Yuan J., Qiang B.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Zheng Q.C., Song Z.F., Zheng Y.W., Li Y.Q., Shu X.;
RT "Molecular cloning and sequencing of human arresten gene."
RL Zhonghua Shi Yan Wai Ke Za Zhi 19:46-47(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Song Z.F., Zheng Q.C.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258349; AA072630.1; -.
DR EMBL; AF363672; AA053382.1; -.
DR EMBL; AF400431; AA092480.1; -.
DR EMBL; AV285780; AA043112.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5C1D5 CRC64;

Query Match 7.0%; Score 17; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 70 VCNFASRNDYSYWLSTP 86

RESULT 21
Q61436
ID Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Muscle;
RC MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal

```

```

DR SMART; SM00111; C4; 2
SQ SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match
Best Local Similarity 7.0%; Score 17; DB 11; Length 585;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 426 VCNFASRNDYSYWLSTP 442
|||||
|||||

RESULT 23
Q8BNS7 PRELIMINARY; PRT; 799 AA.
AC Q8BNS7;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354863; PubMed=12456851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR ENBL; AK080682; BAC37980.1; -.
DR MGD; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR P01413; C4; 2.
DR Pfam; P01391; Collagen; 9.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match
Best Local Similarity 7.0%; Score 17; DB 11; Length 799;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 640 VCNFASRNDYSYWLSTP 656
|||||
|||||

RESULT 24
Q9NUE7 PRELIMINARY; PRT; 886 AA.
AC Q9NUE7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE D242423.1 (collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; C1g helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AB6569 CRC64;

Query Match 7.0%; Score 17; DB 4; Length 886;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 727 VCNFASRNDYSYWLSTP 743
|||||

RESULT 25
Q919X3 PRELIMINARY; PRT; 979 AA.
AC Q919X3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Collagen IV al chain (Fragment).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
basal lamina.";
RL Dev. Biol. 0:0-0(2000).
DR EMBL; AF239838; AAF44681.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 12.
DR ProDom; PD000007; C1g helix; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 7.0%; Score 17; DB 13; Length 979;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 820 VCNFASRNDYSYWLSTP 836
|||||

RESULT 26
Q86X41 PRELIMINARY; PRT; 1075 AA.
ID Q86X41
AC Q86X41;

```

```

DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047305; AAH47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; C1g helix; 3.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match 7.0%; Score 17; DB 4; Length 1075;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 916 VCNFASRNDYSYWLSTP 932
|||||

RESULT 27
Q9H4R9 PRELIMINARY; PRT; 1621 AA.
ID Q9H4R9;
AC Q9H4R9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE BA472K17.2 (Collagen type IV alpha 1 (fragment)).
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390755; CAC13153.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1621 AA; 155705 MW; 73F6F901CDD0EBA2 CRC64;

Query Match 7.0%; Score 17; DB 4; Length 1621;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      84 VCNFASRNDYSYWLSTP 100
Db      1462 VCNFASRNDYSYWLSTP 1478

RESULT 28
QBHYC1
ID      Q8HYC1      PRELIMINARY;      PRT; 1684 AA.
AC      Q8HYC1;
DT      01-MAR-2003 (TREMELrel. 23, Created)
DT      01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT      01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE      Type IV collagen alpha 5 chain (Fragment).
GN      COL4A5.
OS      Canis familiaris (Dog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Harvey S.J., Zheng K., Jefferson B., Sado Y., Naito I., Ninomiya Y.,
RA      Jacobs R., Thorne P.S.;
RT      Recombinant alpha5(IV) collagen: In vivo adenoviral-mediated gene
RT      transfer to smooth muscle restores expression of the alpha6(IV)
RT      collagen chain in a canine model of Alport syndrome."
RL      Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY078501; AAL83712.1; -.
DR      GO; GO:0005581; C:collagen; IEA.
DR      GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR      InterPro; IPR008161; C:helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001442; Procollagn4_C.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 26.
DR      ProDom; PD000007; Clg_helix; 3.
DR      ProDom; PD003923; ProcollagnC4; 1.
DR      SMART; SM00111; C4; 2.
DR      Collagen.
KW      Collagen.
FT      NON_TER
SQ      SEQUENCE 1684 AA; 161408 MW; 02D631B545F285D CRC64;

Query Match      7.0%; Score 17; DB 6; Length 1684;
Best Local Similarity 100.0%; Pred.No. 3.5e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      1532 VCNFASRNDYSYWLSTP 1548

RESULT 29
Q86622
ID      Q86622      PRELIMINARY;      PRT; 1688 AA.
AC      Q86622;
DT      01-JUN-2003 (TREMELrel. 24, Created)
DT      01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT      01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE      Type IV collagen alpha 5.
GN      COL4A5.
OS      Canis familiaris (Dog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Cox M.L., Lees G.E., Kashtan C.E., Murphy K.E.;
RT      "Genetic Cause of X-linked Alport Syndrome in a Family of Domestic
RT      Dogs."
RL      Mamm. Genome 0:0-0(2003).
DR      EMBL; AF470624; AAO33458.1; -.
DR      GO; GO:0005581; C:collagen; IEA.
DR      GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
```

```
DR      InterPro; IPR008161; Clg_helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001442; Procollagn4_C.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 21.
DR      ProDom; PD000007; Clg_helix; 2.
DR      ProDom; PD003923; ProcollagnC4; 2.
DR      SMART; SM00111; C4; 2.
KW      Collagen.
SQ      SEQUENCE 1688 AA; 161725 MW; 7121BE329931CDBC CRC64;

Query Match      7.0%; Score 17; DB 6; Length 1688;
Best Local Similarity 100.0%; Pred.No. 3.5e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      1532 VCNFASRNDYSYWLSTP 1548

RESULT 30
Q8ESQ2
ID      Q8ESQ2      PRELIMINARY;      PRT; 1691 AA.
AC      Q8ESQ2;
DT      01-MAR-2001 (TREMELrel. 16, Created)
DT      01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT      01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE      Type IV collagen alpha 5 chain.
GN      COL4A5.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20536494; PubMed=10965041;
RA      Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,
RA      Kishiro Y., Sado Y., Yoshio H., Ninomiya Y.;
RT      "Differential Expression of Mouse alpha5(IV) and alpha6(IV) Collagen Genes in
RT      Epithelial Basement Membranes."
RL      J. Biochem. 128:427-434(2000).
DR      EMBL; AB041350; BAB13673.1; -.
DR      MGD; MGI:88456; Col4a5.
DR      GO; GO:0005604; C:basement membrane; IEA.
DR      InterPro; IPR008161; Clg_helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001442; Procollagn4_C.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 24.
DR      ProDom; PD000007; Clg_helix; 3.
DR      ProDom; PD003923; ProcollagnC4; 1.
DR      SMART; SM00111; C4; 2.
KW      Collagen.
SQ      SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match      7.0%; Score 17; DB 11; Length 1691;
Best Local Similarity 100.0%; Pred.No. 3.5e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      1532 VCNFASRNDYSYWLSTP 1548

RESULT 31
Q07265
ID      Q07265      PRELIMINARY;      PRT; 1752 AA.
AC      Q07265;
DT      01-NOV-1996 (TREMELrel. 01, Created)
DT      01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT      01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE      3 alpha procollagen.
GN      COL3A1ALPHA.
```

OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8444899;
RX MEDLINE=93186842; Di Liberto M., Ramirez F.;
RA Exposito J.-Y., D'Alessio M., Di Liberto M., Ramirez F.;
RT "Complete primary structure of a sea-urchin type IV collagen and
RT analysis of the 5' end of its gene."
RL J. Biol. Chem. 268:5249-5254(1993).
DR EMBL; L02917; AAA30039.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD000007; C1g_helix; 9.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1752 AA; 170210 MW; 1AESAAR21569346D CRC64;

Query Match 4.9%; Score 12; DB 5; Length 1752;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYSYWLST 99
Db 1597 ASRNDYSYWLST 1608
|||||

RESULT 32
Q9GV24 PRELIMINARY; PRT; 713 AA.
ID Q9GV24
AC Q9GV24
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha-2 (Fragment).
OS Sarcophaga peregrina (Flesh fly) (Boettcherisca peregrina).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
OC Sarcophagidae; Sarcophaga.
OX NCBI_TaxID=7386;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20536508; PubMed=10965055;
RA Fujii-Taira I., Tanaka Y., Homma K.J., Natori S.;
RT "Hydrolysis and synthesis of substrate proteins for cathepsin L in the
RT brain basement membranes of Sarcophaga during metamorphosis."
RL J. Biochem. 128:539-542(2000).
DR EMBL; AS041728; BAB1607.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 713 AA; 74704 MW; 122A361259E40DCB CRC64;

Query Match 4.5%; Score 11; DB 5; Length 713;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLEEFRA 184
|||||

Db 573 SPGSCLEEFRA 583

RESULT 33
O18407 PRELIMINARY; PRT; 1761 AA.
ID O18407
AC O18407
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 2.
GN VKG OR DMCOLA2 OR CG16858.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Yasothornsikul S., Davis W.J., Cramer G., Kimbrell D.A.,
RA Dearolf C.R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55431; AA864082.1; -.
DR FIR; T13990; T13990.
DR FlyBase; FBgn0016075; vkg.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD000007; C1g_helix; 12.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1761 AA; 175955 MW; FCB23AFF19121DC6 CRC64;

Query Match 4.5%; Score 11; DB 5; Length 1761;
Best Local Similarity 100.0%; Pred. No. 0.054;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLEEFRA 184
Db 1668 SPGSCLEEFRA 1678
|||||

RESULT 34
Q9VMV5 PRELIMINARY; PRT; 1940 AA.
ID Q9VMV5
AC Q9VMV5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE VKG protein.
GN VKG OR CG16858.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=Berkeley;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Daluke C., Davenport L.B., Davies P.,
 RA De Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hoscin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moehrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003608; AAF52203.1; -.
 DR FlyBase; Fgn0016075; vkg.
 DR GO; GO:0005587; C:collagen type IV; NAS.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 25.
 DR ProDom; PD000007; Clg_helix; 12.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 SQ SEQUENCE 1940 AA; 193777 MW; 9B507382EF9C17B5 CRC64;

Query Match 4.5%; Score 11; DB 5; Length 1940;
 Best Local Similarity 100.0%; Pred. No. 0.059;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCLEPRA 184
 Db 1670 SPGSCLEPRA 1680

RESULT 35

ID Q14052 PRELIMINARY; PRT; 573 AA.
 AC Q14052;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Alpha-2 type IV collagen (Fragment).
 GN COL4A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88085168; PubMed=3692475;
 RA Killen P.D., Francomano C.A., Yamada Y., Modi W.S., O'Brien S.J.;
 RT "Partial structure of the human alpha 2(IIV) collagen chain and
 RT chromosomal localization of the gene (COL4A2).";
 RL Hum. Genet. 77:318-324(1987).

DR EMBL; M24756; AAA52043.1; -.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; Extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 7.
 DR ProDom; PD000007; Clg_helix; 3.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 FT NON TER
 SQ SEQUENCE 673 AA; 67174 MW; D2F3C9B311A3105 CRC64;

Query Match 4.1%; Score 10; DB 4; Length 673;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAVHSQ 137
 Db 557 PAIAIAVHSQ 566

RESULT 36

ID Q8T7S4 PRELIMINARY; PRT; 1024 AA.
 AC Q8T7S4;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Collagen IV alpha 1 chain (Fragment).
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Anopheles.
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gare D.C., Billingsley P.F.;
 RT "Mosquito collagen IV: conservation of the NC1 domain between alpha
 RT chains.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF133909; AAL99382.1; -.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; Extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 13.
 DR ProDom; PD000007; Clg_helix; 2.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 FT NON TER
 SQ SEQUENCE 1024 AA; 101502 MW; 6BC10D7219C44D68 CRC64;

Query Match 4.1%; Score 10; DB 5; Length 1024;
 Best Local Similarity 100.0%; Pred. No. 0.36;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 PYISRCTVCE 126
 Db 898 PYISRCTVCE 907

RESULT 37

ID Q26640 PRELIMINARY; PRT; 1747 AA.
 AC Q26640;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha2(IV)-like collagen.
GN COLP4ALPHA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=94230414; PubMed=8175744;
RX Exposito J.Y., Suzuki H., Georjon C., Garrone R., Solursh M.,
RA Ramirez F.;
RT "Identification of a cell lineage-specific gene coding for a sea
urchin alpha.2(IV)-like collagen chain.";
RL J. Biol. Chem. 269:13167-13171(1994).
DR EMBL; X76730; CAA54146.1; -.
DR PIR; A54121; A54121.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Collagen.
KW Collagen.
SQ SEQUENCE 1747 AA; 173312 MW; BE722E878394B9B6 CRC64;
Query Match 4.1%; Score 10; DB 5; Length 1747;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 117 PYISRCTVCE 126
Db 1622 PYISRCTVCE 1631
RESULT 38
QBP19 PRELIMINARY; PRT; 59 AA.
ID Q8BP19;
AC Q8BP19;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK075619; BAC35863.1; -.
DR MGD; MGI:88455; Col4a2.
DR GO; GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 1.
DR SMART; SM00111; C4; 1.
DR NON_TER 1
FT SEQUENCE 59 AA; 6697 MW; 61F00BA79B3B4566 CRC64;
Query Match 3.7%; Score 9; DB 11; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ISRCQVCMK 241
Db 49 ISRCQVCMK 57
RESULT 39
Q64457 PRELIMINARY; PRT; 312 AA.
ID Q64457;
AC Q64457;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 4 chain (Fragment).
GN COL4A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Balb/C; TISSUE=Kidney;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sares J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
laminae: Sequence, distribution, association with laminins, and
developmental switches.";
RL J. Cell Biol. 127:879-891(1994).
DR EMBL; Z35167; CAA84530.1; -.
DR PIR; I48303; I48303.
DR MGD; MGI:104687; Col4a4.
DR GO; GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 1.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Collagen.
KW NON_TER 1
FT SEQUENCE 312 AA; 33132 MW; EB017D02868C681E CRC64;
Query Match 3.7%; Score 9; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 233 ISRCQVCMK 241
Db 302 ISRCQVCMK 310
RESULT 40
Q91V13 PRELIMINARY; PRT; 358 AA.
ID Q91V13;
AC Q91V13;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013560; AAH13560.1; -.
DR MGD; MGI:88455; Col4a2.

DR GO: 0005604; C:basement membrane; IDA.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR001442; Procollagn4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 2.
DR ProDom: PD000007; Clg_helix; 1.
DR ProDom: PD003923; ProcollagnC4; 1.
DR SMART: SM00111; C4; 2. Collagen.
KW Hypothetical protein; Collagen.
SQ SEQUENCE 358 AA; 37432 MW; 6DA02BDF3062D9E CRC64;

Query Match 3.7%; Score 9; DB 11; Length 358;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISRCQVCMK 241
Db 348 ISRCQVCMK 356
|||||

RESULT 41
Q9QZR9 PRELIMINARY; PRT; 1682 AA.
AC Q9QZR9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 4 collagen IV.
GN COL4A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC Tissue=Kidney;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL: AF169388; AD50450.1; -.
DR MGD; MGI:104687; Col4a4.
DR GO: 0005604; C:basement membrane; IDA.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR001442; Procollagn4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 22.
DR ProDom: PD000007; Clg_helix; 4.
DR ProDom: PD003923; ProcollagnC4; 1.
DR SMART: SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1682 AA; 164096 MW; 6F7B679EDD76E904 CRC64;

Query Match 3.7%; Score 9; DB 11; Length 1682;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISRCQVCMK 241
Db 1672 ISRCQVCMK 1680
|||||

RESULT 42
Q48996 PRELIMINARY; PRT; 74 AA.
AC Q48996;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE 1-phosphofructinase (Fragment).
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343 (Kid);
RX MEDLINE=96059641; PubMed=7476192;
RA Bork P., Ouzounis C., Casari G., Schneider R., Sander C., Dolan M.,
RA Gilbert W., Gillevet P.M.;
RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
RT its physiology.";
RL Mol. Microbiol. 16:955-967(1995).
DR EMBL: Z33072; CAA83740.1; -.
DR PUR; S77782; S77782.
FT NON_TER 1 74
FT NON_TER 74 74
SQ SEQUENCE 74 AA; 8281 MW; 3AD85E191264D90F CRC64;

Query Match 3.3%; Score 8; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 NYYSNSYS 204
Db 26 NYYSNSYS 33
|||||

RESULT 43
Q8T2W5 PRELIMINARY; PRT; 179 AA.
AC Q8T2W5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TC1114-1.2.
GN TC1114-1.2.
OS Trypanosoma cruzi.
OC Eukaryota; Euzoosoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL Brenner;
RA Anderson B., Bontempi E.J.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC114396; AAL86598.1; -.
DR GO: 0016020; C:membrane; IEA.
DR GO: 0005215; F:transporter activity; IEA.
DR GO: 0006810; P:transport; IEA.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR005829; Sug_transporter.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
SQ SEQUENCE 179 AA; 19990 MW; AE40862E2018E357 CRC64;

Query Match 3.3%; Score 8; DB 5; Length 179;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 FSFLFVQG 52
Db 14 FSFLFVQG 21
|||||

RESULT 44
Q81T03 PRELIMINARY; PRT; 304 AA.
AC Q81T03;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase.
GN BA1489.

| | |
|-----------|---|
| RESULT 47 | |
| Q921S0 | |
| ID | PRELIMINARY; PRT; 452 AA. |
| AC | Q921S0; |
| DT | 01-DEC-2001 (TRENBLrel. 19, Created) |
| DT | 01-DEC-2001 (TRENBLrel. 19, Last sequence update) |
| DT | 01-MAR-2003 (TRENBLrel. 23, Last annotation update) |
| DE | Similar to hypothetical protein FLJ12810. |
| GN | DCLRE1B OR A1452214. |
| OS | Mus musculus (Mouse). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. |
| OX | NCBI_TaxID=10090; |
| RN | [1] |
| PP | SEQUENCE FROM N A |

RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; SC011094; AH011094.1; -;
 DR MGD; MGI:2156057; Dclrelb.
 SQ SEQUENCE 452 AA; 51011 MW; 454DA78A3C235ABD CRC64;

Query Match 3.3%; Score 8; DB 11; Length 452;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDGSGSPA 13
 Db 370 RGDGSGSPA 377

RESULT 48

ID Q8BN95 PRELIMINARY; PRT; 486 AA.
 AC Q8BN95;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE DNA cross-link repair 1B.
 GN DCLRE1B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] -
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Eye;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK084347; BAC39165.1; -;
 DR MGD; MGI:2156057; Dclrelb.
 SQ SEQUENCE 486 AA; 54622 MW; E9F3CFAC1849D3F CRC64;

Query Match 3.3%; Score 8; DB 11; Length 486;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDGSGSPA 13
 Db 404 RGDGSGSPA 411

RESULT 49

ID Q8THZ6 PRELIMINARY; PRT; 523 AA.
 AC Q8THZ6;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein MA4362.
 GN MA4362.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1] -
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McSwan P., McKernan K., Talamas J., Firrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,

RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam I.A., White O., White R.H., de Macario E.C.,
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AB011155; AF007704.1; -;
 DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
 DR GO; GO:0000160; P:two-component signal transduction system (P. . .); IEA.
 DR InterPro; IPR003018; GAF.
 DR InterPro; IPR001610; PAC.
 DR InterPro; IPR000700; PAS-assoc C.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF01590; GAF; 1.
 DR Pfam; PF00785; PAC; 2.
 DR SMART; SM00065; GAF; 1.
 DR SMART; SM00086; PAC; 2.
 DR SMART; SM00091; PAS; 2.
 DR TIGRFAMs; TIGR00229; sensory_box; 2.
 DR PROSITE; PS01113; PAC; 2.
 DR PROSITE; PS01112; PAS; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 523 AA; 59084 MW; C9F2C5CC0AF32C CRC64;

Query Match 3.3%; Score 8; DB 17; Length 523;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 AIAIAVHS 136
 Db 13 AIAIAVHS 20

RESULT 50

Q8C7W7 PRELIMINARY; PRT; 541 AA.
 ID Q8C7W7
 AC Q8C7W7;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE DNA cross-link repair 1B.
 GN DCLRE1B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] -
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK049115; BAC33550.1; -;
 DR MGD; MGI:2156057; Dclrelb.
 SQ SEQUENCE 541 AA; 61055 MW; 281C6AF82F01B76D CRC64;

Query Match 3.3%; Score 8; DB 11; Length 541;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDGSGSPA 13
 Db 459 RGDGSGSPA 466

RESULT 51

Q99K97 PRELIMINARY; PRT; 546 AA.
 ID Q99K97

AC Q99K97; 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN COL4A6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004800; AA04800.1; -.
DR MGD; MGI:2152695; Col4a6.
DR GO; GO:0005587; C:collagen type IV; IDA.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. . .; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Hypothetical protein; Collagen.
FT NON TER 1
SQ SEQUENCE 546 AA; 55102 MW; 56F8CC69374BBCFE CRC64;
Query Match 3.3%; Score 8; DB 11; Length 546;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 234 SRCQVCVK 241
Db 537 SRCQVCVK 544
RESULT 52
Q9C250
ID Q9C250 PRELIMINARY; PRT; 1477 AA.
AC Q9C250;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Related to multidrug resistance-associated protein.
GN B18D24.240.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL513466; CAC28822.2; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001140; ABC TM transport.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 2.

DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW ATP-binding.
SQ SEQUENCE 1477 AA; 161202 MW; 28701BF2084EAB75 CRC64;
Query Match 3.3%; Score 8; DB 3; Length 1477;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 158 IMFTSAGS 165
Db 984 IMFTSAGS 991
RESULT 53
Q9ESQ1
ID Q9ESQ1 PRELIMINARY; PRT; 1691 AA.
AC Q9ESQ1;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Type IV collagen alpha 6 chain.
GN COL4A6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20536494; PubMed=10955041;
RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,
RA Kishiro Y., Sado Y., Yoshioka H., Ninomiya Y.;
RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in
RT Epithelial Basement Membranes."
RL J. Biochem. 128:427-434 (2000).
DR EMBL; AB041351; BAB13674.1; -.
DR MGD; MGI:2152695; Col4a6.
DR GO; GO:0005587; C:collagen type IV; IDA.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. . .; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 22.
DR ProDom; PD000007; Clg_helix; 7.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1691 AA; 164145 MW; CA7E4031DF04F7A7 CRC64;
Query Match 3.3%; Score 8; DB 11; Length 1691;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 234 SRCQVCVK 241
Db 1682 SRCQVCVK 1689
RESULT 54
Q9GQB1
ID Q9GQB1 PRELIMINARY; PRT; 1723 AA.
AC Q9GQB1;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Type IV collagen alpha 1 chain precursor.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564332; PubMed=10956657;

RA Fowler S.J., Jose S., Zhang X., Deutzmann R., Sarraz M.P. Jr.,
RA Boot-Handford R.P., Zhang X., Deutzmann R., Sarraz M.P. Jr.,
RT "Characterization of hydra type IV collagen: Type IV collagen is
RT essential for head regeneration and its expression is up-regulated
RT upon exposure to glucose";
RL J. Biol. Chem. 275:39589-39599 (2000).
DR EMBL; AF282902; AAG40729.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen; Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 1723 AA; 168996 MW; 92496D62FD162F01 CRC64;

Query Match 3.3%; Score 8; DB 5; Length 1723;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTC 196
Db 1669 ECHGRGTC 1676
|||||
RESULT 55
Q9VMV4 PRELIMINARY; PRT; 1779 AA.
ID Q9VMV4
AC Q9VMV4
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CG25C protein.
GN CG25C OR C04145.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
DR EMBL; AF003608; AAF52204.1; -
DR FlyBase; FBgn000299; Cg25C.
DR GO; GO:000587; C:collagen type IV; NAS.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g_helix; 9.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1779 AA; 174300 MW; 6770F18AE40A313B CRC64;

Query Match 3.3%; Score 8; DB 5; Length 1779;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
Db 1769 SRCQVCWK 1776
|||||
RESULT 56
Q9R4U1 PRELIMINARY; PRT; 49 AA.
ID Q9R4U1
AC Q9R4U1
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Iron superoxide dismutase, FE-SOD (EC 1.15.1.1) (Superoxide dismutase
DE [Mn/Fe]) (Fragment).
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
CX NCBI_TaxID=354;
RN [1]
RP SEQUENCE.
RX MEDLINE=95094938; PubMed=8001685;
RA Pagani S., Ciniaghi R., Palagi A., Negri A.;
RT "Purification and characterization of an iron superoxide dismutase
RT from the nitrogen-fixing Azotobacter vinelandii";
RL FEBS Lett. 357:79-82 (1995).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELL AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
DR PIR; S50999; S50999.
DR HSP; P08223; 3SDP.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008901; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom: PD000475; SODismutase; 1.

KW Oxidoreductase.
SQ SEQUENCE 49 AA; 5721 MW; C47F239DCC4E3868 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120

DB 13 ALEPYIS 19

RESULT 57

Q8XW59

ID Q8XW59 PRELIMINARY; PRT; 68 AA.

AC Q8XW59; 2.9%; Score 7; DB 2; Length 49;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein RSC2616.
GN RSC2616 OR RS00918.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502 (2002).
DR EMBL; AL646071; CAD16323.1; -;
KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 68 AA; 7745 MW; 43D0289CF7344211 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 68;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115

DB 50 PITGRAL 56

RESULT 58

Q84ER2

ID Q84ER2 PRELIMINARY; PRT; 69 AA.

AC Q84ER2; 2.9%; Score 7; DB 2; Length 69;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
OS Ralstonia solanacearum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=96344;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A5;
RA Tousseint A.C., Merlin C., Mergay M., Springael D.;
RT "The Biphenyl Catabolic Transposon Tn4371, a Member of a New Family of
RT Genomic Islands Related to IncP and Ti Plasmids";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ536756; CAD61122.1; -;
KW Hypothetical protein.

SQ SEQUENCE 69 AA; 7844 MW; BFDF6917AD4616B2 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115

DB 51 PITGRAL 57

RESULT 59

Q67050

ID Q67050 PRELIMINARY; PRT; 109 AA.

AC Q67050; 2.9%; Score 7; DB 2; Length 109;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Fukuoka/C29/85;
RX MEDLINE=81030852; PubMed=7421990;
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
RT "cloning and dna sequence of double-stranded copies of haemagglutinin
RT genes from h2 and h3 strains elucidates antigenic shift and drift in
RT human influenza virus";
RL Nature 287:301-306 (1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Fukuoka/C29/85;
RX MEDLINE=93323219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza A virus H1 and H2 strains";
RL J. Virol. 67:2552-2558 (1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; D13581; BAA02776.1; -;
DR HSSP; P03437; 1HTM.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001384; Hemagglutn.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171

DB 68 SEGTGQA 74

RESULT 60

Q67053

ID Q67053 PRELIMINARY; PRT; 109 AA.

AC Q67053; 2.9%; Score 7; DB 2; Length 109;
DT 01-NOV-1996 (TREMBlrel. 01, Created)

```
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA
OS Influenzavirus A.
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Sichuan/1/90;
RX MEDLINE=93233219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza A virus H1 and H2 strains.";
RL J. Virol. 67:2552-2558(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC EMBL; D13582; BAA02777.1; -.
DR HSP; P03437; IHGE.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW NON_TER 1
FT NON_TER 109
SQ SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 68 SEGTGQA 74
|||||

RESULT 61
Q67051 PRELIMINARY; PRT; 109 AA.
AC Q67051;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA
OS Influenzavirus A.
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Sichuan/2/87;
RX MEDLINE=93233219; PubMed=7421990;
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
RT "cloning and dna sequence of double-stranded copies of haemagglutinin
RT genes from h2 and h3 strains elucidates antigenic shift and drift in
RT human influenza virus.";
RL Nature 287:301-306(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Ibaraki/1/90;
RX MEDLINE=93233219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza A virus H1 and H2 strains.";
RL J. Virol. 67:2552-2558(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC EMBL; D13583; BAA02778.1; -.
DR HSP; P03437; IHTW.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW NON_TER 1
FT NON_TER 109

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 68 SEGTGQA 74
|||||

RESULT 62
Q67052 PRELIMINARY; PRT; 109 AA.
AC Q67052;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA
OS Influenzavirus A.
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Ibaraki/1/90;
RX MEDLINE=81030852; PubMed=7421990;
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
RT "cloning and dna sequence of double-stranded copies of haemagglutinin
RT genes from h2 and h3 strains elucidates antigenic shift and drift in
RT human influenza virus.";
RL Nature 287:301-306(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Ibaraki/1/90;
RX MEDLINE=93233219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza A virus H1 and H2 strains.";
RL J. Virol. 67:2552-2558(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC EMBL; D13583; BAA02778.1; -.
DR HSP; P03437; IHTW.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW NON_TER 1
FT NON_TER 109
```

```
FT NON TER 109 109
SQ SEQUENCE 109 AA; 12347 MW; 913715121B8F52C6 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTCQA 171
DB 68 SEGTCQA 74
|||||

RESULT 63
Q89PD4 PRELIMINARY; PRT; 116 AA.
AC Q89PD4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BL3548 protein.
GN BL3548.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=2248498; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpoo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005948; BAC48813.1; -.
KW Complete proteome.
SQ SEQUENCE 116 AA; 13293 MW; 21E7C04C4D069A31 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 116;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 EEFASP 186
DB 23 EEFASP 29
|||||

RESULT 64
Q8EKD6 PRELIMINARY; PRT; 118 AA.
AC Q8EKD6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN SO0159.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouiri H., Gill J., Utterback T.R., McDonald L.A.,
```

```
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015466; AAN53246.1; -.
DR TIGR; SC0159; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 118 AA; 13685 MW; 97D1F48EEEC1B45 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 118;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SRNDYSY 95
DB 70 SRNDYSY 76
|||||

RESULT 65
Q9YEN1 PRELIMINARY; PRT; 122 AA.
AC Q9YEN1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0547.
GN APE0547.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KI;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KI.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79515.1; -.
DR PIR; C72639; C72639.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 122 AA; 13447 MW; 08A5685A116D977B CRC64;

Query Match 2.9%; Score 7; DB 17; Length 122;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 DLGTLS 66
DB 52 DLGTLS 58
|||||

RESULT 66
Q68196 PRELIMINARY; PRT; 125 AA.
AC Q68196;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Diol dehydratase-reactivating factor small subunit.
GN DDRB.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
```



```

RC STRAIN=ATCC 8724;
RX MEDLINE=98070363; PubMed=9405397;
RA Mori K., Tobimatsu T., Hara T., Toraya T.;
RT "Characterization, sequencing, and expression of the genes encoding a
RT reactivating factor for glycerol-inactivated adenosylcobalamin-
RL dependent diol dehydratase.";
RL J. Biol. Chem. 272:32034-32041 (1997).
DR EMBL; AF017781; AAC15872.1; -.
DR PIR; T08598; T08598.
SQ SEQUENCE 125 AA; 13620 MW; 8E5E90C59EF74C CRC64;

Query Match      2.9%; Score 7; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAV 134
Db 8 PAIAIAV 14

RESULT 67
ID Q96VT1 PRELIMINARY; PRT; 128 AA.
AC Q96VT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ST2092.
GN ST2092.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140 (2001).
DR EMBL; AP000988; BAB67195.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 128 AA; 14502 MW; E27028A746875C94 CRC64;

Query Match      2.9%; Score 7; DB 17; Length 128;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LKGRGD 8
Db 63 LKGRGD 69

RESULT 68
ID Q9MJR3 PRELIMINARY; PRT; 157 AA.
AC Q9MJR3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NADH dehydrogenase subunit 1 (EC 1.6.5.3) (NADH-ubiquinone
DE oxidoreductase chain 1) (Fragment).
GN NADH1.
OS Taenia pisiformis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Taenia.

```

```

OX NCBI_TaxID=85432;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TP3;
RX MEDLINE=20416034; PubMed=10961852;
RA Gasser R.B., Zhu X., McManus D.P.;
RT "NADH dehydrogenase subunit 1 and cytochrome c oxidase subunit I
RT sequences compared for members of the genus Taenia (Cestoda).";
RL Int. J. Parasitol. 29:1965-1970 (1999).
DR EMBL; AF017781; AAC15872.1; -.
SQ SEQUENCE 125 AA; 13620 MW; 8E5E90C59EF74C CRC64;

Query Match      2.9%; Score 7; DB 8; Length 157;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 YYSNSYS 204
Db 50 YYSNSYS 56

RESULT 69
ID Q9YDQ8 PRELIMINARY; PRT; 168 AA.
AC Q9YDQ8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0859.
GN APE0859.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KJ;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101 (1999).
DR EMBL; AP000060; BAA79839.1; -.
DR PIR; G72679; G72679.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 168 AA; 18599 MW; 267B8179D83E3C6E CRC64;

Query Match      2.9%; Score 7; DB 17; Length 168;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDGSPA 13
Db 113 GDGSPA 119

```

```
RESULT 70
097232 ID 097232 PRELIMINARY; PRT; 171 AA.
AC 097232;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein, PFC0205c (1-cys-glutaredoxin-like
DE protein-1).
GN PFC0205C, MALP2.10 OR GIP-1.
OS Plasmodium falciparum (isolate 3D7), and
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329, 5833;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-P falciparum (isolate 3D7); STRAIN=3D7;
RX MEDLINE=99376085; PubMed=1048855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Hassall B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-P falciparum;
RX MEDLINE=21463082; PubMed=11479312;
RA Rahlfis S., Fischer M., Becker K.;
RT "Plasmodium falciparum Possesses a Classical Glutaredoxin and a
RT Second, Glutaredoxin-like Protein with a PICOT Homology Domain.";
RL J. Biol. Chem. 276:37133-37140 (2001).
DR ENBL; ALQ34558; CAB38997.1;
DR ENBL; AVQ14839; AAK00581.1;
DR GO; GO:0005489; P:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002109; Glutaredoxin.
DR InterPro; IPR004480; Glutaredox-rel.
DR Pfam; PF00462; Glutaredoxin; 1.
DR TIGRFAMs; TIGR00365; TIGR00365; 1.
SQ SEQUENCE 171 AA; 19920 MW; 91E18E5E09E58267 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 171;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 227 GELEKII 233
Db 164 GELEKII 170

RESULT 71
Q8HKQO ID Q8HKQO PRELIMINARY; PRT; 173 AA.
AC Q8HKQO;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 6.
GN ND6.
OS Aspasma minima.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Gobiesocoidae; Gobiesocidae; Aspasma.
OX NCBI_TaxID=181476;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 2.9%; Score 7; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 221 PSTVKAG 227
Db 97 PSTVKAG 103
```

```
RA Miya M.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RA Miya M., Takeshima H., Endo H., Ishiguro N.B., Inoue J.G., Mukai T.,
RA Satoh T.P., Yamaguchi M., Kawaguchi A., Mabuchi K., Shirai S.M.,
RA Nishida M.;
RT "Major Patterns of higher teleostean phylogenies: A new perspective
RT based on 100 complete mitochondrial DNA sequences.";
RL Mol. Phylogenet. Evol. 26:121-138 (2002).
DR ENBL; AP004453; BAC23798.1;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; P:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u...; IEA.
DR InterPro; IPR001457; Oxidored_q3.
DR Pfam; PF00499; oxidored_q3; 1.
DR Mitochondrion;
KW Mitochondrion;
SQ SEQUENCE 173 AA; 17968 MW; B01855C2C1F5124A CRC64;

Query Match 2.9%; Score 7; DB 8; Length 173;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 GTVPLYS 43
Db 132 GTVPLYS 138

RESULT 72
Q82WJO ID Q82WJO PRELIMINARY; PRT; 174 AA.
AC Q82WJO;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Possible cytochrome-c oxidase (EC 1.9.3.1) chain II (EC 1.9.3.1).
GN COXB OR NE0684.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Agciro D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773 (2003).
DR ENBL; BX321858; CAD84595.1;
DR GO; GO:0009481; F:aa3-type cytochrome c oxidase; IEA.
DR GO; GO:0009482; F:ba3-type cytochrome c oxidase; IEA.
DR GO; GO:0009483; F:caa3-type cytochrome c oxidase; IEA.
DR GO; GO:0009485; F:cbb3-type cytochrome c oxidase; IEA.
DR GO; GO:0005507; P:copper ion binding; IEA.
DR GO; GO:0004129; P:cytochrome-c oxidase activity; IEA.
DR InterPro; IPR001505; Copper_CuA.
DR InterPro; IPR008972; Cupredoxin.
DR PROSITE; PS00078; COX2; 1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 174 AA; 19495 MW; 5335A898CC509CDD CRC64;

Query Match 2.9%; Score 7; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 221 PSTVKAG 227
Db 97 PSTVKAG 103
```

```
RESULT 73
Q8SVV5 PRELIMINARY; PRT; 175 AA.
AC Q8SVV5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ECU04_0590.
GN ECU04_0590.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-M1;
RA Genoscope; (APR-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyretallade E., Brotter P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453 (2001).
DR EMBL: AL590444; CAD25245.1; -.
KW Hypothetical protein.
SQ SEQUENCE 175 AA; 20432 MW; 97717F195ADB7D11 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 175;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
DB 149 VPLYSGF 155
|||||

RESULT 74
Q8ND70 PRELIMINARY; PRT; 182 AA.
AC Q8ND70;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN DXF2P667B084.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node;
RA Ausorge W., Winkner U., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL834167; CAD38869.1; -.
DR GO: GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO: GO:0006470; F:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000223; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR PROSITE; PS01032; PP2C; 1.
KW Hypothetical protein.
SQ SEQUENCE 182 AA; 20239 MW; 0D2523DE99A810BB CRC64;

Query Match 2.9%; Score 7; DB 4; Length 182;
```

```
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61
|||||

RESULT 75
Q97ZW3 PRELIMINARY; PRT; 187 AA.
AC Q97ZW3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SSO0461.
GN SSO0461.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-NGOC H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).
DR EMBL: AE006678; AAK40785.1; -.
DR PIR: B90191; B90191.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 187 AA; 21247 MW; 8EC5517CA85D5P22 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 187;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIIS 234
DB 61 ELEKIIIS 67
|||||

RESULT 76
Q59673 PRELIMINARY; PRT; 189 AA.
ID Q59673
AC Q59673;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1) (Fragment).
GN SOD.
OS Propionibacterium freudenreichii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.
OX NCBI_TaxID=1744;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PP3;
RX MEDLINE=95074560; PubMed=7488202;
RA Gabbianelli R., Battistoni A., Polizio F., Carri M.T., De Martino A.,
RA Meier B., Desideri A., Rotilio G.;
RT "Metal uptake of recombinant cambialistic superoxide dismutase from
RT Propionibacterium shermanii is affected by growth conditions of host
RT Escherichia coli cells.";
RL Biochem. Biophys. Res. Commun. 216:841-847 (1995).
RN [2]
RP SEQUENCE FROM N.A.
```

RC STRAIN=P23;
 RA Gabbianelli R.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
 CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
 CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 CC EMBL; X91650; CAA62838.1; -;
 DR EMBL; Y09012; CAA70215.1; -;
 DR HSSP; P80293; LAVM.
 DR GO; GO:0004785; F.copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0006382; F.iron superoxide dismutase activity; IEA.
 DR GO; GO:0008383; F.manganese superoxide dismutase activity; IEA.
 DR GO; GO:0048872; F.metal ion binding; IEA.
 DR GO; GO:0016954; F.nickel superoxide dismutase activity; IEA.
 DR GO; GO:0006801; F.oxidoreductase activity; IEA.
 DR GO; GO:0006801; P.superoxide metabolism; IEA.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sofde; 1.
 DR Pfam; PF02777; sofde_C; 1.
 DR PRINTS; PR01703; MNSODISMUTASE.
 DR PRODOM; PD000475; SODismutase; 1.
 DR PROSITE; PS00088; SOD_MN; 1.
 KW Oxidoreductase.
 FT NON TER 189
 SQ SEQUENCE 189 AA; 21334 MW; 1971419C716F7651 CRC64;

 Query Match 2.9%; Score 7; DB 2; Length 189;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 114 ALPEYIS 120
 DB 15 ALPEYIS 21
 |||||

 RESULT 77
 Q9ZGN1 PRELIMINARY; PRT; 193 AA.
 AC Q9ZGN1
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).
 GN SODB.
 OS Azotobacter vinelandii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Azotobacter.
 OX NCBI_TaxID=354;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CA;
 RA Qurollo B.A., Bishop P.E., Hassan H.M.;
 RT "Identification of the iron superoxide dismutase gene sodB in
 RT Azotobacter vinelandii.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IFO 12018;
 RA Kanematsu S., Sato S., Asada K.;
 RT "Cloning of Azotobacter vinelandii Fe-superoxide dismutase gene and
 RT its anaerobic expression in E. coli cells.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
 CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 CC EMBL; AF077373; AAD02836.1; -;
 DR EMBL; AB025798; BAA88212.1; -;
 DR HSSP; P09223; 3SDP.
 DR GO; GO:0004785; F.copper, zinc superoxide dismutase activity; IEA.

DR GO; GO:0008382; F.iron superoxide dismutase activity; IEA.
 DR GO; GO:0008383; F.manganese superoxide dismutase activity; IEA.
 DR GO; GO:0048872; F.metal ion binding; IEA.
 DR GO; GO:0016954; F.nickel superoxide dismutase activity; IEA.
 DR GO; GO:0006801; F.oxidoreductase activity; IEA.
 DR GO; GO:0006801; P.superoxide metabolism; IEA.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sofde; 1.
 DR Pfam; PF02777; sofde_C; 1.
 DR PRINTS; PR01703; MNSODISMUTASE.
 DR PRODOM; PD000475; SODismutase; 1.
 DR PROSITE; PS00088; SOD_MN; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 193 AA; 21379 MW; 7E879A06D46A422B CRC64;

 Query Match 2.9%; Score 7; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 114 ALPEYIS 120
 DB 14 ALPEYIS 20
 |||||

 RESULT 78
 Q8JRS9 PRELIMINARY; PRT; 199 AA.
 AC Q8JRS9
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Attachment glycoprotein (fragment).
 GN G.
 OS Human respiratory syncytial virus.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
 OX NCBI_TaxID=11250;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=gb427;
 RA Kahn J.S., Martinello R.A., Chen M.D., Weibel C.;
 RT "Correlation between RSV genotype and severity of illness.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF510294; AAM82048.1; -;
 DR InterPro; IPR000925; Glycoprot_G.
 DR Pfam; PF00802; Glycoprotein_G; 1.
 FT NON TER 199
 FT NON TER 199
 SQ SEQUENCE 199 AA; 21724 MW; E74D3189A27DBF74 CRC64;

 Query Match 2.9%; Score 7; DB 12; Length 199;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 26 SCITTAIP 32
 DB 31 SCITTAIP 37
 |||||

 RESULT 79
 Q9A303 PRELIMINARY; PRT; 200 AA.
 AC Q9A303
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cytochrome c oxidase assembly protein, putative.
 GN CC3403.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]

```
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.B., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ernolava M., White O.,
RA Saizberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE006000; AAK25365.1; -.
DR PIR; A87671; A87671.
DR TIGR; CC3403; -.
DR InterPro; IPR007533; CtaG_CoxII.
DR Pfam; PF04442; CtaG_CoxII; 1.
KW Complete proteome.
SQ SEQUENCE 200 AA; 22145 MW; 433F5432810D20AB CRC64;

Query Match 2.9%; Score 7; DB 16; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115
Db 101 PITGRAL 107

RESULT 80
Q28272 PRELIMINARY; PRT; 202 AA.
AC Q28272;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DR GO; GO:0005581; C:collagen; IEA.
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE "Collagen type IV alpha 2 chain (Fragment)."
GN COL4A2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation."
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50934; AAC48594.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4_C4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
FT NON TER 202
FT NON TER 202
SQ SEQUENCE 202 AA; 22079 MW; 25A56E7642A329FC CRC64;

Query Match 2.9%; Score 7; DB 6; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AIAVHSQ 137
Db 105 AIAVHSQ 111

RESULT 81
Q83HP6
```

```
ID Q83HP6 PRELIMINARY; PRT; 202 AA.
AC Q83HP6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1).
GN SODA OR TW473.
OS Tropheryma whippelii (strain TW08/27) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=218496;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22495039; PubMed=12606174;
RA Bentley S.D., Maiwald M., Murphy L.D., Pallen M.J., Yeats C.A.,
RA Dover L.G., Norbertczak H.T., Besta G.S., Quail M.A., Harris D.E.,
RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,
RA Barrall B.G., Parkhill J., Relman D.A.;
RT "Sequencing and analysis of the genome of the Whipple's disease
RT bacterium Tropheryma whippelii.";
RL Lancet 361:637-644(2003).
DR EMBL; BX251411; CAD67140.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR ProDom; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD MN; 1.
DR Oxidoreductase; Complete proteome.
KW SEQUENCE 202 AA; 22900 MW; 15305AA77FD42964 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 15 ALEPYIS 21

RESULT 82
Q83G14 PRELIMINARY; PRT; 202 AA.
AC Q83G14;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1).
GN SODA OR TW299.
OS Tropheryma whippelii (strain Twist) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=203267;
RN [1]
RP SEQUENCE FROM N.A.
RA Raoult D., Audic S., Robert C., Ogata H., Suhr K., Drancourt M.,
RA Claverie J.-M.;
RT "Tropheryma whippelii illustrates the diversity of gene loss patterns
RT in small genome bacterial pathogens."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016851; AAC44396.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
```

DR GO:0016491; F:oxidoreductase activity; IEA.
 DR GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro: IPR001189; SODismutase.
 DR Pfam: PF00081; sodfe; 1.
 DR Pfam: PF02777; sodfe C; 1.
 DR PRINTS: PR01703; MNSODISMUTASE.
 DR ProDom: PD000475; SODismutase; 1.
 DR ProSITE: PS00088; SOD_MN; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 202 AA; 22901 MW; 9738FEEB999EBD28 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 202;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
 |||||
 Db 15 ALEPYIS 21

RESULT 83

Q28274 PRELIMINARY; PRT; 205 AA.

AC Q28274;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Collagen type IV alpha 5 chain (Fragment).
 GN COL4A6
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Samoyed;
 RX MEDLINE=96278820; PubMed=8662866;
 RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
 RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
 of collagen type IV. Evidence from a canine model of X-linked
 nephritis with a COL4A5 gene mutation."
 RL J. Biol. Chem. 271:13821-13828(1996).
 DR EMBL: J50937; AAC48587.1; -.
 DR GO:0005581; C:collagen; IEA.
 DR GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro: IPR001442; Procollagn4_C.
 DR Pfam: PF01413; C4; 2.
 DR ProDom: PD003923; ProcollagnC4; 1.
 DR SMART: SM00111; C4; 2.
 FT NON_TER 1
 FT NON_TER 205
 SQ SEQUENCE 205 AA; 22708 MW; 84D7BDE4C1C00395 CRC64;

Query Match 2.9%; Score 7; DB 6; Length 205;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AIAVHSQ 137
 |||||
 Db 107 AIAVHSQ 113

RESULT 84

Q96UT6 PRELIMINARY; PRT; 206 AA.

AC Q96UT6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Manganese-containing superoxide dismutase (EC 1.15.1.1) (Superoxide
 dismutase [Mn/Fe]).
 GN SOD3.
 OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamarre C., LeMay J.-D., Deslauriers N., Bourbonnais Y.;
 RT "Candida albicans expresses an unusual cytoplasmic manganese-
 containing superoxide dismutase SOD3 upon entry and during stationary
 phase."
 RL J. Biol. Chem. 276:20000-20006(2001).
 CC -|- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
 CC -|- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
 CC -|- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 FAMILY.
 CC EMBL: AF416340; AAL08560.1; -.
 DR GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO:0008382; F:iron superoxide dismutase activity; IEA.
 DR GO:0008383; F:manganese superoxide dismutase activity; IEA.
 DR GO:0046872; F:metal ion binding; IEA.
 DR GO:0016954; F:nickel superoxide dismutase activity; IEA.
 DR GO:0016491; F:oxidoreductase activity; IEA.
 DR GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro: IPR001189; SODismutase.
 DR Pfam: PF00081; sodfe; 1.
 DR PRINTS: PR01703; MNSODISMUTASE.
 DR ProDom: PD000475; SODismutase; 1.
 DR ProSITE: PS00088; SOD_MN; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 206 AA; 22734 MW; 303FA3503417F332 CRC64;

Query Match 2.9%; Score 7; DB 3; Length 206;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
 |||||
 Db 19 ALEPYIS 25

RESULT 85

Q29468 PRELIMINARY; PRT; 208 AA.

AC Q29468;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Collagen type IV alpha 4 chain (Fragment).
 GN COL4A4.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Samoyed;
 RX MEDLINE=96278820; PubMed=8662866;
 RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
 RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
 of collagen type IV. Evidence from a canine model of X-linked
 nephritis with a COL4A5 gene mutation."
 RL J. Biol. Chem. 271:13821-13828(1996).
 DR EMBL: U50936; AAC48586.1; -.
 DR GO:0005581; C:collagen; IEA.
 DR GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro: IPR001442; Procollagn4_C.
 DR Pfam: PF01413; C4; 2.
 DR ProDom: PD003923; ProcollagnC4; 1.
 DR SMART: SM00111; C4; 2.
 FT NON_TER 1
 FT NON_TER 208
 SQ SEQUENCE 208 AA; 23135 MW; 136077AB651A21FC CRC64;

```

Query Match          2.9%; Score 7; DB 6; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
DB 148 SPGSCLE 154

RESULT 86
Q96Y84 PRELIMINARY; PRT; 211 AA.
AC Q96Y84;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative superoxide dismutase.
GN ST2283.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=2145156; PubMed=11572479;
RA Kawaiabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000989; BAB67393.1; -.
DR GO; GO:0046872; P:metal ion binding; IEA.
DR GO; GO:0004784; P:superoxide dismutase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; scdfe; 1.
DR Pfam; PF02777; scdfe; 1.
DR PRINTS; PRO1703; MNSODISMUTASE.
DR ProDom; PD000475; SODismutase; 1.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 211 AA; 24302 MW; 0EEAC4AA76938002 CRC64;

Query Match          2.9%; Score 7; DB 17; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPVIS 120
DB 21 ALEPVIS 27

RESULT 87
P95356 PRELIMINARY; PRT; 216 AA.
AC P95356;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE FtsE-like protein.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH811;
RX MEDLINE=20277473; PubMed=10819322;

```

```

RA Bernatchez S., Francis P.M., Salimma H., Beveridge T.J., Li H.,
RA Dillon J.-A.R.;
RT "Genomic, transcriptional and phenotypic analysis of ftsE and ftsX of
RT Neisseria gonorrhoeae.";
RL DNA Res. 7:175-81(2000).
RL -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; U76418; AAB36524.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005286; IISF.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00960; 3a0501s02; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 216 AA; 23906 MW; 125F8E5363167FEB CRC64;

Query Match          2.9%; Score 7; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GDLGLTL 64
DB 63 GDLGLTL 69

RESULT 88
Q9KLR3 PRELIMINARY; PRT; 216 AA.
AC Q9KLR3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell division ATP-binding protein FtsE.
GN NMB0007.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
RC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC_TRANSPORTERS)
DR EMBL; AE002359; AAP40486.1; -.
DR PIR; E81247; E81247.
DR TIGR; NMB0007; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.

```

```
DR InterPro: IPR005286; IISP.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRFAWS: TIGR00960; 3a0501s02; 1.
DR TRITRAN: TRITRAN001; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_2; 1.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Complete Proteome.
SQ SEQUENCE 216 AA; 23945 MW; 11572764FC125FEC CRC64;

Query Match          2.9%; Score 7; DB 16; Length 216;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GQDLGTL 64
   |||||
DB 63 GQDLGTL 69

RESULT 89
Q9JW1 PRELIMINARY; PRT; 216 AA.
AC Q9JW1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE ABC transporter (TrEMBLrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein.
GN FTSE OR NMA0254.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4a;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagals K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC_TRANSPORTERS).
DR EMBL; AL162752; CAB83562.1; -.
DR PIR; B82020; B82020.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR005286; IISF.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRFAWS: TIGR00960; 3a0501s02; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Complete Proteome.
SQ SEQUENCE 216 AA; 23961 MW; 53F6E07B8B904DE2 CRC64;

Query Match          2.9%; Score 7; DB 16; Length 216;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GQDLGTL 64
   |||||
DB 63 GQDLGTL 69
```

```
RESULT 90
Q8MU16 PRELIMINARY; PRT; 220 AA.
AC Q8MU16;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mn superoxide dismutase (EC 1.15.1.1) (Superoxide dismutase
DE [MnFe]).
OS Trichinella pseudospiralis.
OC Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida;
OC Trichinellidae; Trichinella.
OX NCBI_TaxID=6337;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu H.W.K., Ko R., Mak C.H.;
RT "Characterization and molecular cloning of Mn- superoxide dismutase in
RT Trichinella pseudospiralis."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
DR EMBL; AF521909; AAM76074.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro: IPR001189; SODismutase.
DR Pfam: PF00081; sodfe; 1.
DR Pfam: PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR ProDom: PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase.
SQ SEQUENCE 220 AA; 24574 MW; 85BA976A1P367FF7 CRC64;

Query Match          2.9%; Score 7; DB 5; Length 220;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
   |||||
DB 32 ALEPYIS 38

RESULT 91
Q8IXG7 PRELIMINARY; PRT; 233 AA.
AC Q8IXG7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UG0882E07.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y.;
RT "Isolation of full-length cDNA clones from human fetal brain cDNA
RT library."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF351614; AAN76514.1; -.
DR PIR; PT0240; PT0240.
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
```


DR GO: GO:0003824; F:catalytic activity; IEA.
 DR GO: GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
 DR GO: GO:0005470; P:protein amino acid dephosphorylation; IEA.
 DR InterPro: IPR000222; PP2C.
 DR InterPro: IPR001932; PP2C-like.
 DR Pfam: PF00461; PP2C; 1.
 DR PROSITE: PS01032; PP2C; 1.
 SQ SEQUENCE 233 AA; 25652 MW; EB90A7B3BC1BDD08 CRC64;

Query Match 2.9%; Score 7; DB 4; Length 233;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
 DB 55 SGSPATW 61

RESULT 92

Q7V4W7 PRELIMINARY; PRT; 236 AA.

ID Q7V4W7;
 AC Q7V4W7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN PMT1820.
 OS Prochlorococcus marinus (strain MIT 9313).
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=74547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22825698; PubMed=12917642;
 RA Roca G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
 RA Angren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
 RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
 RA Shaw S.L., Stegler C., Sullivan M.B., Ting C.S., Tolonen A.,
 RA Webb E.A., Zinser E.R., Chisholm S.W.;
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
 niche differentiation."
 RL Nature 424:1042-1047(2003).
 DR EMBL; BX572100; CAE21995.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 236 AA; 27651 MW; 5964906532FEB20E CRC64;

Query Match 2.9%; Score 7; DB 16; Length 236;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 FLFCNVN 82
 DB 83 FLFCNVN 89

RESULT 93

Q9K8B8 PRELIMINARY; PRT; 237 AA.

ID Q9K8B8;
 AC Q9K8B8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein BH3088.
 GN BH3088.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11059132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AP001517; BAB06907.1; -.
 DR PIR; H84035; H84035.
 DR InterPro: IPR008535; DUF817.
 DR Pfam; PF05675; DUF817; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 237 AA; 27822 MW; 2F2EB9A29BF14B30 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 237;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
 DB 100 VPLYSGF 106

RESULT 94

Q8F8X2 PRELIMINARY; PRT; 239 AA.

ID Q8F8X2;
 AC Q8F8X2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN LA0427.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 OX NCBI_TaxID=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
 RA Ren S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE011230; AAN47626.1; -.
 DR InterPro: IPR005660; DUF344.
 DR Pfam; PF03976; DUF344; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 239 AA; 28774 MW; ECB100FBFB7900F CRC64;

Query Match 2.9%; Score 7; DB 16; Length 239;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIIS 234
 DB 233 ELEKIIIS 239

RESULT 95

Q8ML71 PRELIMINARY; PRT; 241 AA.

ID Q8ML71;
 AC Q8ML71;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE CG18279-PB.
 GN CG18279.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Change M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhattacharya D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman I.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwac C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-F., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celisner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwac M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RL "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminck J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celisner S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Venter C.J., Lewis S.E.;
 RL "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003818; AAM68571.1;
 DR FlyBase; FBgn0033835; CG18279.

SQ SEQUENCE 241 AA; 25652 MW; EAEEA311134F9E1B CRC64;
 Query Match 2.9%; Score 7; DB 5; Length 241;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 IPSCPEG 37
 Db 217 IPSCPEG 223
 RESULT 96
 Q8N2M3 PRELIMINARY; PRT; 244 AA.
 ID Q8N2M3
 AC Q8N2M3; 2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein FLJ90120.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
 RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK074601; BAC11085.1;
 KW Hypothetical protein.
 SQ SEQUENCE 244 AA; 24989 MW; F11285DC202EFC43 CRC64;
 Query Match 2.9%; Score 7; DB 4; Length 244;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 171 ALASPGS 177
 Db 21 ALASPGS 27
 RESULT 97
 Q8G7Z6 PRELIMINARY; PRT; 245 AA.
 ID Q8G7Z6
 AC Q8G7Z6; 2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Narrowly conserved hypothetical transmembrane protein.
 CN BL0088
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 CX NCBI_TaxID=216816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwielen M.-C., Desiere F., Bork P., Delley M.,
 RA Pridmore R.D., Arigoni F.;
 RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL; AE014623; AAN23953.1;
 DR GO; GO:0016021; C: integral to membrane; IEA.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 SQ SEQUENCE 245 AA; 26057 MW; E6BA8DE8DB3EC48E CRC64;

```
Query Match          2.9%; Score 7; DB 16; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 93 YSYWLST 99
    143 YSYWLST 149
    |||||
    |||||

RESULT 98
QXEX8
ID QXEX8 PRELIMINARY; PRT; 247 AA.
AC QXEX8;
DT 01-WAR-2002 (TrEMBLrel. 20, Created)
DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein yacF (putative cytoplasmic protein) (Hypothetical
DE protein STY0161).
GN YACF OR T0145 OR STM0139 OR STY0161.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601, 602;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700933;
RX MEDLINE=22531357; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McLelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor T.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR ENBL; AEO16834; AAO67877.1; -
DR ENBL; AEO08700; AAL19103.1; -
DR ENBL; AL627265; CAD01298.1; -
KW Hypothetical protein; complete proteome.
SQ SEQUENCE 247 AA; 28425 MW; E1B9826AD004B548 CRC64;

Query Match          2.9%; Score 7; DB 16; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 WLASLNP 212
    206 WLASLNP 212
    |||||
    |||||
```

```
Db 157 WLASLNP 163

RESULT 99
Q87BW1
ID Q87BW1 PRELIMINARY; PRT; 252 AA.
AC Q87BW1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 3-deoxy-manno-octulosonate cytidyltransferase.
GN KDSB OR PD1337.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Moon D.H.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., da Silva F.R.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Xuranae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Secubal J.C.,
RA Kitajima J.P.;
RA "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RN fastidiosa."
RL J. Bacteriol. 185:1018-1026(2003).
DR ENBL; AEO12558; AAO29184.1; -
DR GO; GO:0016779; Fnucoacyldltransferase activity; IEA.
DR GO; GO:0016740; Ftransferase activity; IEA.
DR GO; GO:0009103; Polypolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR Pfam; PF02348; CTP transf.3; 1
KW Transferase; Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 252 AA; 27400 MW; 1D6B0A3FFE231619 CRC64;

Query Match          2.9%; Score 7; DB 16; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRFTTM 74
    194 LQRFTTM 200
    |||||
    |||||

Db 194 LQRFTTM 200

RESULT 100
Q81YK7
ID Q81YK7 PRELIMINARY; PRT; 254 AA.
AC Q81YK7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Membrane protein, putative.
GN BA3535.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
```

RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
 RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
 RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
 RA Thonason B., Friedlander A.M., Koshler T.M., Hanna P.C., Kolsto A.-B.,
 RA Fraser C.M.;
 RT "The genome sequence of *Bacillus anthracis* Ames and comparison to
 RT closely related bacteria.";
 RL NATURE 423:81-86(2003).
 DR EMBL; AB017035; AAP27298.1; --
 DR TIGR; BA3535; --
 KW Complete proteome.
 SQ SEQUENCE 254 AA; 30326 MW; 5D1554DA1E41329A CRC64;

Query Match 2.9%; Score 7; DB 16; Length 254;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
 DB 103 VPLYSGF 109

RESULT 101

Q81AT7 PRELIMINARY; PRT; 254 AA.

AC Q81AT7;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN BC3469.
 OS *Bacillus cereus* (strain ATCC 14579 / DSM 31).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=226900;
 RN [1]
 RX SEQUENCE FROM N.A.
 RX MEDLINE=22608415; PubMed=12721630;
 RA Ivancova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
 RA Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
 RA Chu L., Mazur N., Goltsman B., Larsen N., D'Souza M., Walunas T.,
 RA Grechkin Y., Pusch G., Haseikorn R., Feinstein M., Ehrlich S.D.,
 RA Overbeek R., Kyrpides N.;
 RT "Genome sequence of *Bacillus cereus* and comparative analysis with
 RT *Bacillus anthracis*.";
 RL NATURE 423:87-91(2003).
 DR EMBL; AB017009; AAP10404.1; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 254 AA; 30257 MW; 0016C115197A3D1 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 254;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
 DB 103 VPLYSGF 109

RESULT 102

Q92WN8 PRELIMINARY; PRT; 255 AA.

AC Q92WN8;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Iron-superoxide dismutase (EC 1.15.1.1) (Superoxide dismutase
 DE [Mn/Fe]).
 OS *Oryza sativa* (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzeae; Oryza.

OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=cv. Nipponbare;
 RX MEDLINE=99208990; PubMed=10192910;
 RA Kaminaka H., Morita S., Tokumoto M., Yokoyama H., Masumura T.,
 RA Tanaka K.;
 RA "Molecular cloning and characterization of a cDNA for an iron-
 RT superoxide dismutase in rice (*Oryza sativa* L.).";
 RL Biosci. Biotechnol. Biochem. 63:302-308(1999).
 CC -I- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
 CC -I- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
 CC -I- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 DR EMBL; AB014056; BAA37131.1; --
 DR PIR; JG0179; JG0179.
 DR HSP; P09223; 3SDP.
 DR Gramene; Q92WN8; --
 DR GO; GO:0004785; F.copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0008382; F.iron superoxide dismutase activity; IEA.
 DR GO; GO:0008383; F.manganese superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F.metal ion binding; IEA.
 DR GO; GO:0016954; F.nickel superoxide dismutase activity; IEA.
 DR GO; GO:0016491; F.oxidoreductase activity; IEA.
 DR GO; GO:0006801; P.superoxide metabolism; IEA.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sofde_1.
 DR Pfam; PF02777; sofde_C_1.
 DR PRINTS; PR01703; MNSODISMUTASE.
 DR PRODOM; PD000475; SODismutase; 1.
 DR PROSITE; PS00088; SOD_MN; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 255 AA; 29369 MW; 2DE803FA6161B443 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 255;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPFPYIS 120
 DB 54 ALPFPYIS 60

RESULT 103

Q9PB46 PRELIMINARY; PRT; 257 AA.

AC Q9PB46;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE 3-deoxy-manno-octulosonate cytidyltransferase.
 GN XF2299.
 OS *Xylella fastidiosa*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorzy H.,
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Praga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.P.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.F.O., Netto L.E.S.,
 RA Nnanl A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Tenenly M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Vertovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
 RL Nature 406:151-159(2000).
 DR EMBL; AF004041; AAF85098.1; --
 DR PIR; H82575; H82575.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0008690; F:3-deoxy-manno-octulosonate cytidyltransferase; IEA.
 DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
 DR InterPro; IPR003329; Cytidylyl_trans.
 DR InterPro; IPR004528; KdsB.
 DR Pfam; PF02348; CTP_transf_3; 1.
 DR TIGRFAMs; TIGR00486; kdsB; 1.
 KW Complete proteome.
 SQ SEQUENCE 257 AA; 27883 MW; 6CBB08A5D78B0BC8 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 257;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LQRTTMM 74
 Db 199 LQRTTMM 205

RESULT 104

Q8SM64 PRELIMINARY; PRT; 259 AA.
 AC Q8SM64;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Superoxide dismutase 3 (Fragment).
 OS Toxoplasma gondii.
 OG Mitochondrion.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
 OC Toxoplasma.
 OX NCBI_TaxID=5811;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ding M., Kwok L.Y., Krauth-Siegel L., Schlueter D., Clayton C.,
 RA Soldati D.;
 RT "Role of catalase and peroxiredoxins as defence mechanism against
 RT oxidative stress in *Toxoplasma gondii*.";
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AIZ54045; AAC72229.1; --
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0004682; F:metal ion binding; IEA.
 DR GO; GO:0004784; F:superoxide dismutase activity; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe; 1.
 DR Pfam; PF02777; sodfe_C; 1.
 DR PRINTS; PR01703; MNSODISMUTASE.
 DR ProDom; PD000475; SODismutase; 1.
 KW Mitochondrion.
 FT NON_TER 259
 SQ SEQUENCE 259 AA; 23864 MW; C0C3D8D0D6F2B6E5 CRC64;

Query Match 2.9%; Score 7; DB 8; Length 259;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALBPYIS 120
 Db 63 ALBPYIS 69

RESULT 105

Q9SNQ0 PRELIMINARY; PRT; 259 AA.
 ID Q9SNQ0;
 AC Q9SNQ0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ESTs C26547 (C12563) (EC 1.15.1.1) (Superoxide dismutase
 DE [Mn/Fel])
 OS *Oryza sativa* (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
 RT clone:POS35G04.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
 CC - SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 DR EMBL; AP000399; BAA83577.1; --
 DR HSSP; P09223; 3SDP.
 DR Gramene; Q9SNQ0; --
 DR GO; GO:004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
 DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe; 1.
 DR Pfam; PF02777; sodfe_C; 1.
 DR PRINTS; PR01703; MNSODISMUTASE.
 DR ProDom; PD000475; SODismutase; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 259 AA; 29905 MW; F685EC8AC3C15A88 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 259;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALBPYIS 120
 Db 54 ALBPYIS 60

RESULT 106

Q34535 PRELIMINARY; PRT; 264 AA.
 ID Q34535;
 AC Q34535;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE YOAT.
 GN YOAT.
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]

```

RP SEQUENCE FROM N.A.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich D.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertoletto M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Boursier L., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrai E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF027868; AAB84424.1; -.
DR ENBL; Z99114; CAB13767.1; -.
DR PIR; E69897; E69897.
DR InterPro; IPR008535; DUF817.
DR Pfam; PF05675; DUF817; 1.
KW Complete proteome.
SQ SEQUENCE 264 AA; 30605 MW; D31FD23AB2E8F667 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 VPLYSGF 45
Db 103 VPLYSGF 109

RESULT 107
Q92C22 PRELIMINARY; PRT; 267 AA.
ID Q92C22
AC Q92C22;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lin1029.
GN LIN1029.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
CX NCBI_TaxID=1642;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR ENBL; AL59167; CAC96260.1; -.
DR PIR; AD1561; AD1561.
DR Listalist; LIN01029; -.
DR InterPro; IPR008535; DUF817.
DR Pfam; PF05675; DUF817; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 267 AA; 31450 MW; 7C15IC69740DB8C2 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 VPLYSGF 45
Db 103 VPLYSGF 109

RESULT 108
Q8Y880 PRELIMINARY; PRT; 267 AA.
ID Q8Y880
AC Q8Y880;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lmo1037.
GN LMO1037.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
CX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR ENBL; AL59167; CAC99115.1; -.
DR PIR; AE1204; AE1204.
DR Listalist; LMO01037; -.
DR InterPro; IPR008535; DUF817.
DR Pfam; PF05675; DUF817; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 267 AA; 31204 MW; EAA24A6DEB6266D0 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 39 VPLXSGF 45
 |||||
 Db 103 VPLXSGF 109

RESULT 109

Q8ET45 PRELIMINARY; PRT; 267 AA.
 AC Q8ET45;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical conserved protein.
 GN OB0418.
 OS Oceanobacillus iheyensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HT831 / DSM 14371 / JCM 11309;
 RX MEDLINE=2220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments.";
 RL Nucleic Acids Res. 30:3927-3935 (2002).
 DR EMBL; AF004594; BAC12374.1; -;
 DR InterPro; IPR008535; DUF817.
 DR Pfam; PF05675; DUF817; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 267 AA; 31291 MW; A45A3EC074FC41E3 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 267;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLXSGF 45
 |||||
 Db 103 VPLXSGF 109

RESULT 110

Q7V2M5 PRELIMINARY; PRT; 274 AA.
 AC Q7V2M5;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative multidrug efflux ABC transporter.
 GN PM00450.
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=59919;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22825698; PubMed=12917642;
 RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
 RA Ahlstrom N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
 RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
 RA Webb E.A., Zinser E.R., Chisholm S.W.;
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
 RT niche differentiation."
 RL Nature 424:1042-1047 (2003).
 DR EMBL; BX572091; CAE18909.1; -;
 KW Complete proteome.
 SQ SEQUENCE 274 AA; 30471 MW; 0FOAACD4806F3BB9 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 274;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212
 |||||
 Db 209 WLASLNP 215

RESULT 111

Q9LWS3 PRELIMINARY; PRT; 282 AA.
 AC Q9LWS3;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Similar to Zantedeschia aethiopica iron superoxide dismutase
 DE (EC 1.15.1.1) (Superoxide dismutase [Mn/Fe]).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
 RT clone: P0541H01.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
 CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 DR EMBL; AP001389; BAA92737.1; -;
 DR HSSP; PI9665; IQNN.
 DR Gramene; Q9LWS3; -;
 DR GO; GO:0004785; F.copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0008382; F.iron superoxide dismutase activity; IEA.
 DR GO; GO:0008383; F.manganese superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F.metal ion binding; IEA.
 DR GO; GO:0016954; F.nickel superoxide dismutase activity; IEA.
 DR GO; GO:0016491; F.oxidoreductase activity; IEA.
 DR GO; GO:0006801; F.superoxide metabolism; IEA.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe; 1.
 DR Pfam; PF02777; sodfe C; 1.
 DR PRINTS; PR01703; MNSODISMUTASE.
 DR ProDom; PD000475; SODismutase; 1;
 DR Oxidoreductase.
 KW Oxidoreductase.
 SQ SEQUENCE 282 AA; 32124 MW; 6E0A907B3732CA52 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 282;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
 |||||
 Db 35 ALEPYIS 41

RESULT 112

Q8Q083 PRELIMINARY; PRT; 285 AA.
 AC Q8Q083;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase subunit
 DE X.
 GN MM0254.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=Goel / G01 / ATCC BAR-199 / DSM 3647 / OCM 88;
RX  MEDLINE=22120827; PubMed=12125824;
RA  Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA  Martinez-Arias R., Henne A., Wieser A., Baeumer S., Jacobi C.,
RA  Brueggemann H., Llenard T., Christmann A., Boencke M., Steckel S.,
RA  Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA  Fritz H.-J., Gottschalk G.;
RT  "The genome of Mechanosarcina mazzei: evidence for lateral gene
RT  transfer between Bacteria and Archaea.";
RL  J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR  EMBL; AB013250; AM29950.1; -.
DR  GO; GO:0008168; F:mechyltransferase activity; IEA.
DR  GO; GO:0016740; F:transferase activity; IEA.
KW  Transferase; Methyltransferase; Complete proteome.
SQ  SEQUENCE 285 AA; 30743 MW; 55BCD6753ADB44C2 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 285;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 ITGRALE 116
Db 190 ITGRALE 196
|||||

RESULT 113
Q8F0K1 PRELIMINARY; PRT; 297 AA.
AC Q8F0K1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN LA3492.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=56601 / Serogroup Icterohaemorrhagiae / Seroovar lai;
RX MEDLINE=56601; PubMed=1107;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS011506; AN50690.1; -.
DR InterPro; IPR001107; Band 7.
DR Pfam; PF01145; Band 7; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 297 AA; 34391 MW; FE14184912AEAE83 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 STVAGE 228
Db 48 STVAGE 54
|||||

RESULT 114
Q88GA3 PRELIMINARY; PRT; 304 AA.
AC Q88GA3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome c family protein.
GN PF3822.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=22423050; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Medler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Dueterhoeft A., Tuemmier B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AF016788; AA69416.1; -.
DR TIGR; P3822; -.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR00345; CytC_heme_BS.
DR InterPro; IPR003088; Cyt_C1.
DR Pfam; PF00034; cytochrome c; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 2.
KW Complete proteome.
SQ SEQUENCE 304 AA; 33251 MW; DSA3EF698F1B716 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 93 YSYWLST 99
Db 147 YSYWLST 153
|||||

RESULT 115
Q8C7L9 PRELIMINARY; PRT; 313 AA.
AC Q8C7L9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DnaJ (Fragment).
GN DNAJ6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=hippocampus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK049935; BAC3992.1; -.
DR PIR; PT0635; PT0696.
DR MGD; MGI:1919935; Dnajc6.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DNAJ_2; 1.
FT NON_TER 1
FT 1
SQ SEQUENCE 313 AA; 33978 MW; 98C5E25DB98B957F CRC64;

Query Match 2.9%; Score 7; DB 11; Length 313;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 DLGTLS 66
|||||

```


Db 81 DLGTLGS 87

RESULT 116

Q82ZY2 PRELIMINARY; PRT; 313 AA.
 AC Q82ZY2;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Binding protein, putative.
 GN PAB0020.
 OS Pyrobaculum aerophilum.
 CC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
 CC Thermoproteaceae; Pyrobaculum.
 CX NCBI_TaxID=13773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
 RX MEDLINE=21664397; PubMed=11792869;
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
 RA Miller J.H.;
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
 aerophilum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
 DR EMBL; AE009746; AAL62507.1; -;
 KW Complete proteome.
 SQ SEQUENCE 313 AA; 34199 MW; E754B5E72B6CA7C4 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 313;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 GTGQALA 173

Db 90 GTGQALA 96

RESULT 117

Q98L58 PRELIMINARY; PRT; 316 AA.
 AC Q98L58;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Probable mdcF malonate transporter.
 GN MLL1169.
 OS Rhizobium loti (Mesorhizobium loti).
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Phyllobacteriaceae; Mesorhizobium.
 CX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shingo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP002996; BAB48605.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR InterPro; IPR004776; Auxin_eff.
 DR DR Pfam; PF03547; Auxin_eff; I.
 KW Complete proteome.
 SQ SEQUENCE 316 AA; 33254 MW; 329010C7D28EA1F6 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 316;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TQALAS 174
 Db 284 TQALAS 290

RESULT 118

Q88B27 PRELIMINARY; PRT; 331 AA.
 AC Q88B27;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN PSP00205.
 OS Pseudomonas syringae (pv. tomato).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 CX NCBI_TaxID=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000.
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
 RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
 RA White O., Fraser C., Collier A.;
 RT "Complete sequence of Pseudomonas syringae.";
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AS016856; AAO53752.1; -;
 DR TIGR; PSP00205; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 331 AA; 36137 MW; A4DD78A55437C197 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 331;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SGFSFLF 49

Db 13 SGFSFLF 19

RESULT 119

Q92T66 PRELIMINARY; PRT; 338 AA.
 AC Q92T66;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative oxidoreductase protein.
 GN R00114 OR SMC04138.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 CX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rameberger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galbert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591782; CAC41501.1; -;
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000683; GFO_IDH_MOCA.
 DR InterPro; IPR004104; GFO_IDH_MOCA_C.

DR Pfam; PF01408; GFO_IDH_MocA; 1.
DR Pfam; PF02894; GFO_IDH_MocA_C; 1.
KW Complete proteome.
SQ SEQUENCE 338 AA; 37006 MW; EBEFFBE47EACD0B99 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 TVKAGEL 229
DB 132 TVKAGEL 138
|||||
RESULT 120
Q9SAC7 PRELIMINARY; PRT; 343 AA.
ID Q9SAC7
AC Q9SAC7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE T16B5.11 protein.
GN T16B5.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=sv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J., Liu A., Li J., Kremenetskaia I., Luros J., Gonzalez A.,
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC T16B5 sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007354; RAD31338.1; -.
DR PIR; A66241; A66241.
DR InterPro; IPR004314; DUF239.
DR Pfam; PF03080; DUF239; 1.
SQ SEQUENCE 343 AA; 38362 MW; 441E8F36C5D324CF CRC64;

Query Match 2.9%; Score 7; DB 10; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CPEGTV 40
DB 55 CPEGTV 61
|||||
RESULT 121
Q915Y7 PRELIMINARY; PRT; 354 AA.
ID Q915Y7
AC Q915Y7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PA0549.
GN PA0549.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004491; AAG03938.1; -.
DR PIR; C83577; C83577.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 354 AA; 41087 MW; ASB55A1E79008F3E CRC64;

Query Match 2.9%; Score 7; DB 16; Length 354;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
DB 115 SPGSCLE 121
|||||
RESULT 122
O56962 PRELIMINARY; PRT; 365 AA.
ID O56962
AC O56962;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Haemagglutinin (Fragment).
GN HA
OS Influenza A virus (A/equine/Newmarket-Bob Champion/89(H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=71445;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=A/equine/Newmarket-Bob Champion/89(H3N8);
RX MEDLINE=98309086; PubMed=9645136;
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ223193; CAAL1168.1; -.
DR HSSP; P03437; 2VIU.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT CHAIN 1
FT NON_TER 1
FT CHAIN 1
FT NON_TER 1
FT CHAIN 330
FT NON_TER 365
SQ SEQUENCE 365 AA; 40393 MW; 59ED645AB56485B1 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 358 SEGTGOA 364
|||||
RESULT 123
O56961

```

ID O56961 PRELIMINARY; PRT; 365 AA.
AC O56961;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Haemagglutinin (Fragment).
GN HA.
OS Influenza A virus (A/equine/Lichfield/89(H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=71440;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A /equine/Lichfield/89(H3N8);
RX MEDLINE=98309066; PubMed=9645196;
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ223192; CAA11167.1; -.
DR HSPF; P03437; 2VIU.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 1
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >365 HA2 HAEMAGGLUTININ.
FT NON TER 365
FT NON TER 365
SQ SEQUENCE 365 AA; 40403 MW; E47DD6931659861 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGSGQA 171
Db 358 SEGSGQA 364
|||||

RESULT 124
ID Q9DL24 PRELIMINARY; PRT; 371 AA.
AC Q9DL24;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Athens/135/99 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=147343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Athens/135/99;
RA Plakoukafalos E.T., Markoulatos P., Spyrou N., Vamvakopoulos N.;
RT "Molecular and phylogenetic analysis of hemagglutinin and
RT neuraminidase sequences from recent human influenza type A (H3N2)
RT viral isolates in Southern Greece.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

ID O56961 PRELIMINARY; PRT; 365 AA.
AC O56961;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Haemagglutinin (Fragment).
GN HA.
OS Influenza A virus (A/equine/Lichfield/89(H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=71440;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A /equine/Lichfield/89(H3N8);
RX MEDLINE=98309066; PubMed=9645196;
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ223192; CAA11167.1; -.
DR HSPF; P03437; 2VIU.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 1
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >365 HA2 HAEMAGGLUTININ.
FT NON TER 365
FT NON TER 365
SQ SEQUENCE 365 AA; 40403 MW; E47DD6931659861 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGSGQA 171
Db 358 SEGSGQA 364
|||||

RESULT 124
ID Q9DL24 PRELIMINARY; PRT; 371 AA.
AC Q9DL24;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Athens/135/99 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=147343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Athens/135/99;
RA Plakoukafalos E.T., Markoulatos P., Spyrou N., Vamvakopoulos N.;
RT "Molecular and phylogenetic analysis of hemagglutinin and
RT neuraminidase sequences from recent human influenza type A (H3N2)
RT viral isolates in Southern Greece.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

ID O56961 PRELIMINARY; PRT; 365 AA.
AC O56961;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Haemagglutinin (Fragment).
GN HA.
OS Influenza A virus (A/equine/Lichfield/89(H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=71440;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A /equine/Lichfield/89(H3N8);
RX MEDLINE=98309066; PubMed=9645196;
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ223192; CAA11167.1; -.
DR HSPF; P03437; 2VIU.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 1
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >365 HA2 HAEMAGGLUTININ.
FT NON TER 365
FT NON TER 365
SQ SEQUENCE 365 AA; 40403 MW; E47DD6931659861 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGSGQA 171
Db 365 SEGSGQA 371
|||||

RESULT 125
Q96NT4 PRELIMINARY; PRT; 372 AA.
ID Q96NT4;
AC Q96NT4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ30116.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Negahari K., Masuko Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE PP2C FAMILY.
DR EMBL; AK054678; BAB70790.1; -.
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000222; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00352; PP2Cc; 1.
DR SMART; SM00351; PP2C-SIG; 1.
DR PROSITE; PS01032; PP2C; 1.
DR Hypothetical protein; Hydrolase; Magnesium.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE 372 AA; 40983 MW; 7065B29DC79CB93B CRC64;

Query Match 2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15
Db 55 SGSPATW 61
|||||

RESULT 126
```

```
Q8N3J5          PRELIMINARY;          PRT;      372 AA.
ID Q8N3J5
AC Q8N3J5
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN DAFZP761G058. (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Amalgam;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL34271; CAD38946.1; -.
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000222; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00332; PP2CC; 1.
DR SMART; SM00331; PP2C SIG; 1.
DR PROSITE; PS01032; PP2C; 1.
KW Hypothetical protein.
KW SEQUENCE 372 AA; 40997 MW; 9DD37ECC0EAD3313 CRC64;

Query Match      2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 SGSPATW 15
DB      55 SGSPATW 61

RESULT 127
Q8IUZ7          PRELIMINARY;          PRT;      372 AA.
ID Q8IUZ7
AC Q8IUZ7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancras;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037552; AAH37552.1; -.
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000222; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00332; PP2CC; 1.
DR SMART; SM00331; PP2C SIG; 1.
DR PROSITE; PS01032; PP2C; 1.
KW Hypothetical protein.
KW SEQUENCE 372 AA; 41011 MW; 29927CB2BDD32A2 CRC64;

Query Match      2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 SGSPATW 15
DB      55 SGSPATW 61

RESULTS 128
Q8H666          PRELIMINARY;          PRT;      382 AA.
ID Q8H666
AC Q8H666
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative iron superoxide dismutase.
GN OSUNBA0019F11.12.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Embryophyta; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC
clone:CSUNBA0019F11.1";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002837; BAC22204.1; -.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0004784; F:superoxide dismutase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
SQ SEQUENCE 382 AA; 42202 MW; D0C871B0FD0E3BDE CRC64;

Query Match      2.9%; Score 7; DB 10; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      114 ALEPYIS 120
DB      135 ALEPYIS 141

RESULTS 129
Q8UK63          PRELIMINARY;          PRT;      384 AA.
ID Q8UK63
AC Q8UK63
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN H3HA.
OS Influenza A virus (A/teal/Germany/wv201r/01).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=205472;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/teal/Germany/wv01r/01;
RA Werner O., Starick E., Mueller T., Muehle R.;
RT "Characterisation of avian influenza virus isolates from wild birds
from Germany";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
```

```
DR EMBL: AJ506781; CAD44999.1; -.
DR GO: GO:0019321; C:Viral envelope; IEA.
DR InterPro: IPR008980; Capsid_hemag.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 384 384
SQ SEQUENCE 384 AA; 42076 MW; 459731795CA5C838 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 384;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 374 SEGTGQA 380
|||||

RESULT 130
Q7XSV3 PRELIMINARY; PRT; 389 AA.
AC Q7XSV3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBAA039K24.3 protein.
GN OSJNBAA039K24.3
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Wang Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.X., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.B., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (Sep-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL606637; CAB01784.1; -.
SQ SEQUENCE 389 AA; 42397 MW; 554D9E0BD4590A02 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 CLEEFRA 184
Db 94 CLEEFRA 100
|||||

RESULT 131
Q89C31 PRELIMINARY; PRT; 398 AA.
AC Q89C31;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BIL7967 protein.
GN BIL7967.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.

STRAIN-USDA 110;
MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shingo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005964; BAC53232.1; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR005625; DUF337.
DR InterPro: IPR000362; Fumarate_lyase.
DR Pfam: PF03929; DUF337; 1.
DR PROSITE; PS00163; FUMARATE_LYASES; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 43470 MW; E143B907A7D6EA36 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 398;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 STPALMP 104
Db 59 STPALMP 65
|||||

RESULT 132
Q9NWF7 PRELIMINARY; PRT; 403 AA.
AC Q9NWF7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K.,
RA Arata M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000919; BAA91424.1; -.
DR InterPro: IPR000313; PWWP_domain.
DR Pfam: PF00855; PWWP; 1.
DR PROSITE; PS50812; PWWP; 1.
KW Hypothetical protein.
SQ SEQUENCE 403 AA; 45754 MW; 1B5D2124275EF437 CRC64;

Query Match 2.9%; Score 7; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
Db 328 SPGSCLE 334
|||||

RESULT 133
Q8EG37 PRELIMINARY; PRT; 407 AA.
AC Q8EG37;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
```

GN SOL1774.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.C., Venter J.C., Neallson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR ENBL; A2015621; AAN54827.1; -;
DR TIGR; SOL1774; -;
DR InterPro; IPR008599; Diacid rec.
DR Pfam; PF05651; Diacid rec; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 407 AA; 45442 MW; 533DD15CE0F210C CRC64;

Query Match 2.9%; Score 7; DB 16; Length 407;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GQALSP 175
DB 323 GQALSP 329
|||||
RESULT 134
Q9Q0L5 PRELIMINARY; PRT; 409 AA.
AC Q9Q0L5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Swine/North Carolina/35922/98 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=101753;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Swine/North Carolina/35922/98;
RX MEDLINE=99412408; PubMed=10482643;
RA Zhou N.N., Senne D.A., Landgraf J.S., Swenson S.L., Erickson G.,
RA Rossow K., Liu L., Yoon K.J., Krause S., Webster R.G.;
RT "Genetic reassortment of avian, swine, and human influenza A viruses
RT in American pigs";
RL J. Virol. 73:8851-8856 (1999).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR ENBL; AF153232; AAD51239.1; -;
DR HSP; P03437; 2VIU.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 409
SQ SEQUENCE 409 AA; 45296 MW; 9866D50F46E531F2 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 409;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 374 SEGTGQA 380
|||||
RESULT 135
Q93SI5 PRELIMINARY; PRT; 412 AA.
AC Q93SI5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE O-antigen acetylase WbIA.
GN WbIA.
OS Burkholderia thailandensis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=57975;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99141012; PubMed=9989483;
RA Deshazer D., Brett P.J., Woods D.E.;
RT "The type II O-antigenic polysaccharide moiety of Burkholderia
RT pseudomallei lipopolysaccharide is required for serum resistance and
RT virulence";
RL Mol. Microbiol. 30:1081-1100 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Brett P.J., Burtick M.N., Woods D.E.;
RT "WbIA Activity is Required for the 2-O-Acetylation of O-Antigens
RT Expressed by Burkholderia pseudomallei and Burkholderia
RT thailandensis";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY033501; AAK51152.1; -;
DR GO; GO:0016747; F:transferase activity, transferring groups o. . .; IEA.
DR InterPro; IPR002656; Acyl_transf_3.
DR Pfam; PF01757; Acyl_transf_3; 1.
SQ SEQUENCE 412 AA; 45630 MW; 301FD0F7538549EF CRC64;

Query Match 2.9%; Score 7; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SRNDYSY 95
DB 331 SRNDYSY 337
|||||
RESULT 136
ID O69122 PRELIMINARY; PRT; 412 AA.
AC O69122;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative O-antigen acetylase.
GN WbIA.
OS Burkholderia pseudomallei (Pseudomonas pseudomallei), and
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=28450; 13373;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.pseudomallei; STRAIN=1026B;
RA Deshazer D., Brett P.J., Woods D.E.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RC SPECIES=B.mallei;
RA Burtnick M.N., Brett P.J., Woods D.E.;
RT "Physical and Molecular Characterization of Lipopolysaccharide O-
antigens from Burkholderia mallei";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF064070; AAD05460.1; -;
DR EMBL: AY028370; AAK27397.1; -;
DR GO: GO:0016747; F:transferase activity, transferring groups o. . .; IEA.
DR InterPro: IPR002656; Acyl_transf_3.
DR Pfam: PF01757; Acyl_transf_3; 1.
SQ SEQUENCE 412 AA; 45682 MW; DDDF409C70BF9747 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 SRNDYSY 95
Db 331 SRNDYSY 337
|||||

RESULT 137
Q8KN89 ID Q8KN89 PRELIMINARY; PRT; 416 AA.
AC Q8KN89;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF_8.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kuttyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
aeruginosa";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL: AF498407; AAM27645.1; -;
SQ SEQUENCE 416 AA; 46332 MW; E1661BB8F401319B CRC64;

Query Match 2.9%; Score 7; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 ISLWKG 155
Db 295 ISLWKG 301
|||||

RESULT 138
Q8KN80 ID Q8KN80 PRELIMINARY; PRT; 416 AA.
AC Q8KN80;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF_8.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kuttyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
aeruginosa";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL: AF498407; AAM27645.1; -;
SQ SEQUENCE 416 AA; 46332 MW; E1661BB8F401319B CRC64;

RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL: AF498414; AAM27766.1; -;
SQ SEQUENCE 416 AA; 46346 MW; D6646B03603536C9 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 ISLWKG 155
Db 295 ISLWKG 301
|||||

RESULT 139
Q9QJ4 ID Q9QJ4 PRELIMINARY; PRT; 416 AA.
AC Q9QJ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Swine/Iowa/8548-1/98).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=101751;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=A/Swine/Iowa/8548-1/98;
RX MEDLINE=99412408; PubMed=10482643;
RA Zhou N.N., Senne D.A., Landgraf J.S., Swenson S.L., Erickson G.,
RA Rossow K., Liu L., Yoon K.J., Krauss S., Webster R.G.;
RT "Genetic reassortment of avian, swine, and human influenza A viruses
in American pigs";
RL J. Virol. 73:8851-8856(1999).
CC - FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: AF153235; AAD51242.1; -;
DR HSP; P03437; 2VIU.
DR GO: GO:0019031; C: viral envelope; IEA.
DR InterPro: IPR009980; Capsid hemag.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 416
SQ SEQUENCE 416 AA; 46324 MW; 753E943B03CAB0ED CRC64;

Query Match 2.9%; Score 7; DB 12; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 374 SEGTGQA 380
|||||

RESULT 140
Q92043 ID Q92043 PRELIMINARY; PRT; 419 AA.
AC Q92043;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemorrhagic toxin a (EC 3.4.24.1) (Fragment).
GN HT-A, ATROLYSIN A.
OS Crocotalus atrox (Western diamondback rattlesnake).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Crotalus.

```
OX NCBI_TaxID=8730;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=venom gland;
RA Hite L.A., Jia L.-G., Bjarnason J.B., Fox J.W.;
RT "cDNA sequences for four snake venom metalloproteinases: Structure,
RT classification, and their relationship to mammalian reproductive
RT proteins.";
RL Arch. Biochem. Biophys. 0:0-0(1993).
DR EMBL; U01234; AAA0326.1; -.
DR PIR; S41607; S41607.
DR HSP; P15167; IDTH.
DR MEROPS; M12.142; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00200; disintegrin_1.
DR Pfam; PF01421; Reprolysin_1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SMC0608; ACR; 1.
DR SMART; SMC0050; DISIN; 1.
DR PROSITE; PS02015; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS02014; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase.
FT NON_TER
SQ SEQUENCE 419 AA; 46879 MW; 442833518478E416 CRC64;

Query Match 2.9%; Score 7; DB 13; Length 419;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 77 LFCNVND 83
DB 367 LFCNVND 373

RESULT 141
OY Q9GSD3 PRELIMINARY; PRT; 424 AA.
AC Q9GSD3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative transporter protein CG10.
GN CG10.
OS Plasmodium vivax (strain Salvador I).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=126793;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Salvador I;
RX MEDLINE=21240730; PubMed=11343215;
RA Nomura T., Carlton J.M.R., Baird J.K., Del Portillo H.A.,
RA Fryauff D.J., Rathore D., Fidock D.A., Su X.-Z., Collins W.E.,
RA McCutchan T.F., Wootton J.C., Wellens T.E.;
RT "Evidence for different mechanisms of chloroquine resistance in 2
RT Plasmodium species that cause human malaria.";
RL J. Infect. Dis. 183:1653-1661(2001).
DR EMBL; AF314649; AAG27739.1; -.
DR EMBL; AF314649; AAG27739.1; -.
SQ SEQUENCE 424 AA; 48569 MW; EB54A287C401BFD9 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY Q9NH61 PRELIMINARY; PRT; 424 AA.
AC Q9NH61
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

```
OY 127 GPAIAIA 133
DB 352 GPAIAIA 358

RESULT 142
OY Q8T9R7 PRELIMINARY; PRT; 424 AA.
AC Q8T9R7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative chloroquine resistance transporter.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM6;
RA Li G.-D., Ward S.A.;
RT "Plasmodium falciparum TM6 putative chloroquine resistance transporter
RT (crt) mRNA.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF468066; AAL75580.1; -.
DR EMBL; AF468066; AAL75580.1; -.
SQ SEQUENCE 424 AA; 48581 MW; DE470716070B49D6 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 GPAIAIA 133
DB 353 GPAIAIA 359

RESULT 143
OY Q9GSD7 PRELIMINARY; PRT; 424 AA.
AC Q9GSD7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative transporter protein CG10.
GN CG10.
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21240730; PubMed=11343215;
RA Nomura T., Carlton J.M.R., Baird J.K., Del Portillo H.A.,
RA Fryauff D.J., Rathore D., Fidock D.A., Su X.-Z., Collins W.E.,
RA McCutchan T.F., Wootton J.C., Wellens T.E.;
RT "Evidence for different mechanisms of chloroquine resistance in 2
RT Plasmodium species that cause human malaria.";
RL J. Infect. Dis. 183:1653-1661(2001).
DR EMBL; AF314646; AAG27735.1; -.
DR EMBL; AF314646; AAG27735.1; -.
SQ SEQUENCE 424 AA; 48672 MW; E3F2872D08988753 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 GPAIAIA 133
DB 352 GPAIAIA 358

RESULT 144
OY Q9NH61 PRELIMINARY; PRT; 424 AA.
AC Q9NH61
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```



```
QY 127 GPATAIA 133
DB 353 GPATAIA 359

RESULT 149
Q9ST79 PRELIMINARY; PRT; 425 AA.
AC Q9ST79;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CAA303718.1 protein.
GN Q3037.18.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DNA;
RA Hong G., Chen Z.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ245900; CAB53491.1; -.
DR Gramene; Q9ST79; -.
DR InterPro; IPR008511; DUF793.
DR Pfam; PF05633; DUF793; 1.
SQ SEQUENCE 425 AA; 46455 MW; C6D3ERC56DC4B777 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 425;
Best Local Similarity 100.0%; Pred.No. 2.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 CLEPRA 184
DB 94 CLEPRA 100

RESULT 150
Q9Q0L4 PRELIMINARY; PRT; 429 AA.
AC Q9Q0L4;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Swine/Texas/4199-2/98(H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=136487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Swine/Texas/4199-2/98;
RX MEDLINE=9412408; PubMed=10482643;
RA Zhou N.N., Senne D.A., Landgraf J.S., Swenson S.L., Erickson G.,
RA Rossow K., Liu L., Yoon K.J., Krauss S., Webster R.G.;
RT "Genetic reassortment of avian, swine, and human influenza A viruses
in American pigs.";
RL J. Virol. 73:8851-8856(1999).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DE EMBL; AF153233; AAD51240.1; -.
DR HSP; P03437; 1HTM.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
```

```
DR ProDom: PD000225; Hemagglutin; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 429 429
SQ SEQUENCE 429 AA; 47838 MW; 72DC7925F8D02285 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 429;
Best Local Similarity 100.0%; Pred.No. 2.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTOQA 171
DB 374 SEGTOQA 380

Search completed: April 5, 2004, 07:38:57
Job time : 55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:05:43 ; Search time 60 Seconds

(without alignments)
1149.027 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRGDSGPATWTRGF.....KAGLEKIISRCQVCKKEH 244

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database :

A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 244 | 100.0 | 244 | 5 | ABG79219 Human Goo |
| 2 | 244 | 100.0 | 244 | 5 | Aau75595 Human typ |
| 3 | 244 | 100.0 | 244 | 6 | ADA20225 Human typ |
| 4 | 244 | 100.0 | 245 | 3 | Aay67942 Human typ |
| 5 | 244 | 100.0 | 245 | 5 | Aau75589 Human typ |
| 6 | 244 | 100.0 | 1670 | 7 | ADA47063 Human Pro |
| 7 | 235 | 96.3 | 244 | 5 | ABG79218 Human typ |
| 8 | 235 | 96.3 | 244 | 5 | ABG79217 Human typ |
| 9 | 191 | 78.3 | 191 | 5 | Aau75596 Human typ |
| 10 | 191 | 78.3 | 191 | 6 | ADA20260 Human typ |
| 11 | 183 | 75.0 | 254 | 5 | Aau75598 Human typ |
| 12 | 163 | 66.8 | 268 | 2 | Aay31993 Human typ |
| 13 | 163 | 66.8 | 268 | 3 | Aay97555 Human alp |
| 14 | 159 | 65.2 | 211 | 3 | Aay95918 Human Goo |
| 15 | 159 | 65.2 | 211 | 5 | ABG79208 Human GP |
| 16 | 155 | 63.5 | 232 | 7 | ADCI17697 Human typ |
| 17 | 141 | 57.8 | 218 | 2 | Aay44172 Human typ |
| 18 | 141 | 57.8 | 218 | 3 | Aay56784 Human alp |
| 19 | 141 | 57.8 | 218 | 4 | Aae09484 Human alp |
| 20 | 132 | 54.1 | 132 | 6 | ADA20261 Human typ |
| 21 | 131 | 53.7 | 132 | 5 | Aau75597 Human typ |
| 22 | 124 | 50.8 | 124 | 5 | Aau75594 Human typ |
| 23 | 124 | 50.8 | 124 | 6 | ADA20258 Human typ |
| 24 | 120 | 49.2 | 120 | 6 | ADA20259 Human typ |
| 25 | 112 | 45.9 | 112 | 6 | ADA20262 Human typ |

| | | | | | | |
|----|----|------|------|---|-----------|-----------|
| 26 | 99 | 40.6 | 218 | 2 | AAU792164 | Partial s |
| 27 | 88 | 36.1 | 88 | 5 | AAU75607 | Human typ |
| 28 | 88 | 36.1 | 88 | 6 | ADA20271 | Human typ |
| 29 | 81 | 33.2 | 88 | 5 | AAU75608 | Human typ |
| 30 | 80 | 32.8 | 88 | 6 | ADA20272 | Human typ |
| 31 | 79 | 32.4 | 79 | 5 | AAU75600 | Human typ |
| 32 | 79 | 32.4 | 79 | 6 | ADA20264 | Human typ |
| 33 | 65 | 26.6 | 65 | 5 | AAU75599 | Human typ |
| 34 | 64 | 26.2 | 64 | 6 | ADA20263 | Human typ |
| 35 | 61 | 25.0 | 68 | 3 | AAU95920 | Human Goo |
| 36 | 61 | 25.0 | 68 | 5 | ABG79210 | Human GP |
| 37 | 61 | 25.0 | 72 | 3 | AAU95919 | Human Goo |
| 38 | 61 | 25.0 | 72 | 3 | AAU95921 | Human Goo |
| 39 | 61 | 25.0 | 72 | 5 | ABG79209 | Human GP |
| 40 | 61 | 25.0 | 72 | 5 | ABG79211 | Human GP |
| 41 | 39 | 16.0 | 230 | 7 | ADA47061 | Rat Prote |
| 42 | 39 | 16.0 | 471 | 2 | AAU44171 | Bovine ty |
| 43 | 39 | 16.0 | 471 | 3 | AAU56783 | Bovine al |
| 44 | 39 | 16.0 | 471 | 4 | AAE09483 | Bovine al |
| 45 | 37 | 15.2 | 72 | 5 | ABG79213 | Human GP |
| 46 | 36 | 14.8 | 36 | 4 | AAE09503 | Human C8 |
| 47 | 30 | 12.3 | 471 | 2 | AAU792163 | Partial s |
| 48 | 26 | 10.7 | 26 | 4 | AAE09501 | Human C7 |
| 49 | 26 | 10.7 | 27 | 6 | ADA20238 | T8 peptid |
| 50 | 25 | 10.2 | 25 | 6 | ADA20236 | T7 peptid |
| 51 | 22 | 9.0 | 22 | 7 | ADCI17661 | Type IV c |
| 52 | 22 | 9.0 | 22 | 7 | ADCI17414 | Type IV c |
| 53 | 21 | 8.6 | 21 | 3 | AAU95912 | Human Goo |
| 54 | 21 | 8.6 | 21 | 5 | ABG79202 | Human Goo |
| 55 | 21 | 8.6 | 21 | 7 | ADCI17642 | Type IV c |
| 56 | 20 | 8.2 | 20 | 5 | AAU75604 | Human typ |
| 57 | 20 | 8.2 | 20 | 5 | AAU75602 | Human typ |
| 58 | 20 | 8.2 | 20 | 5 | AAU75603 | Human typ |
| 59 | 20 | 8.2 | 20 | 6 | ADA20267 | Human typ |
| 60 | 20 | 8.2 | 20 | 6 | ADA20266 | Human typ |
| 61 | 20 | 8.2 | 20 | 6 | ADA20268 | Human typ |
| 62 | 20 | 8.2 | 20 | 7 | ADCI17684 | Type IV c |
| 63 | 19 | 7.8 | 19 | 5 | AAU75606 | Human typ |
| 64 | 19 | 7.8 | 19 | 5 | AAU75605 | Human typ |
| 65 | 19 | 7.8 | 19 | 6 | ABP58053 | Collagen |
| 66 | 19 | 7.8 | 19 | 6 | ADA20269 | Human typ |
| 67 | 19 | 7.8 | 19 | 6 | ADA20265 | Human typ |
| 68 | 19 | 7.8 | 19 | 6 | ADA20270 | Human typ |
| 69 | 19 | 7.8 | 19 | 7 | AAU75601 | Human typ |
| 70 | 18 | 7.4 | 18 | 7 | ADCI17672 | Type IV c |
| 71 | 18 | 7.4 | 18 | 7 | ADCI17672 | Type IV c |
| 72 | 18 | 7.4 | 18 | 7 | ADCI17649 | Type IV c |
| 73 | 17 | 7.0 | 46 | 4 | AAU18657 | Peptide # |
| 74 | 17 | 7.0 | 46 | 5 | ABG40558 | Human pep |
| 75 | 17 | 7.0 | 229 | 1 | AAU95524 | Complete |
| 76 | 17 | 7.0 | 229 | 3 | AAU67943 | Human typ |
| 77 | 17 | 7.0 | 229 | 5 | AAU75587 | Human typ |
| 78 | 17 | 7.0 | 229 | 6 | ADA20217 | Human typ |
| 79 | 17 | 7.0 | 229 | 7 | ADCI17695 | Human typ |
| 80 | 17 | 7.0 | 229 | 7 | ADCI17699 | Human typ |
| 81 | 17 | 7.0 | 260 | 2 | AAU31991 | Type IV c |
| 82 | 17 | 7.0 | 260 | 3 | AAU97553 | Human alp |
| 83 | 17 | 7.0 | 264 | 2 | AAU31995 | Type IV c |
| 84 | 17 | 7.0 | 264 | 3 | AAU97557 | Human alp |
| 85 | 17 | 7.0 | 309 | 3 | AAU54044 | Human pan |
| 86 | 17 | 7.0 | 406 | 3 | AAU58169 | Lung canc |
| 87 | 17 | 7.0 | 772 | 2 | AAU23873 | Human alp |
| 88 | 17 | 7.0 | 772 | 2 | AAU09643 | Human typ |
| 89 | 17 | 7.0 | 1669 | 4 | AAU40863 | Human pol |
| 90 | 17 | 7.0 | 1669 | 5 | ABG90760 | Human Tum |
| 91 | 17 | 7.0 | 1669 | 6 | ABU57334 | Mouse isc |
| 92 | 17 | 7.0 | 1669 | 6 | ABU54467 | Human tum |
| 93 | 17 | 7.0 | 1672 | 4 | AAU39077 | Human hum |
| 94 | 17 | 7.0 | 1685 | 4 | ABG04839 | Novel hum |
| 95 | 17 | 7.0 | 1693 | 4 | ABG15619 | Novel hum |
| 96 | 16 | 6.6 | 16 | 7 | ADCI17470 | Type IV c |
| 97 | 15 | 6.1 | 15 | 4 | AAE09491 | Human C2 |
| 98 | 15 | 6.1 | 15 | 4 | AAE09497 | Human C5 |

PS Disclosure; Page 213-214; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an

CC autoimmune condition by identifying compounds that reduce phosphorylation

CC of a first target protein (I) (which is selected from Goodpasture antigen

CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)

CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-

CC Ala-Thr-Thr-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising

CC Glu-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of

CC conformational isomers of the second target protein (II) (selected from

CC an alpha3 type IV collagen NC1 domain polypeptide and myelin basic

CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3

CC NC1 domain conformational isomer, which has an amino acid sequence

CC identical to the wild type alpha3 type IV collagen NC1 domain, is

CC stabilised by disulphide bonds, and has a molecular weight in a non-

CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in

CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated

CC type IV collagen alpha3 NC1 domain. The human gene for GPBP is located on

CC chromosome 5q13. The method is useful for treating autoimmune conditions,

CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous

CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present

CC sequence represents a Goodpasture syndrome related protein or peptide

XX

SQ Sequence 244 AA;

Query Match 100.0%; Score 244; DB 5; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.2e-241;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGSGSPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60

DB 1 GLKGRGSGSPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60

QY 61 LGTLGSLQRFMTTPELFCNVNDVCFNFRNDYSYWLSTPALMPMNPATITGRALEPYIS 120

DB 61 LGTLGSLQRFMTTPELFCNVNDVCFNFRNDYSYWLSTPALMPMNPATITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

QY 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIESTVKAGELEKIIISRCQVCM 240

DB 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIESTVKAGELEKIIISRCQVCM 240

QY 241 KKRH 244

DB 241 KKRH 244

RESULT 2

AAU75595

ID AAU75595 standard; protein; 244 AA.

XX AAU75595;

XX

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin 334.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;

XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX

PR 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-189037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

XX treating disorders involving angiogenesis.

XX Example 33; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

XX domain, having one or more of the characteristics selected from: (a) the

XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit

XX proliferation of endothelial cells; and (c) the ability to cause

XX apoptosis of endothelial cells. Also described are the following: (1) use

XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

XX analogue or allelic variant in the preparation of a medicament for

XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

XX where the angiogenesis is mediated by one or more endothelial cell

XX integrins or one or more endothelial cell integrin subunits; or (b) by

XX promoting or inducing endothelial cell apoptosis in a tissue, where the

XX endothelial cell apoptosis is mediated by one or more endothelial cell

XX integrins or one or more endothelial cell integrin subunits; (2) use of

XX an antibody or peptide that specifically binds the alpha1, alpha2, or

XX alpha3, alpha4, alpha5, alpha6, beta1 or beta2 subunit of integrin in the

XX preparation of a medicament for inhibiting angiogenesis or cell

XX proliferation; (3) use of an inhibitor, such as an antibody, antibody

XX fragment or peptide of receptor-mediated angiogenesis in a vertebrate,

XX where the disease is characterised by angiogenesis that is mediated by

XX receptors to Arresten, Canstatin or Tumstatin and where the receptors

XX inhibited are Arresten, Canstatin or Tumstatin; (4) use of one

XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in

XX the presence of a medicament for promoting angiogenesis in a tissue; and

XX (5) use of integrins in the preparation of a medicament for promoting or

XX inducing angiogenesis or cell proliferation in a tissue. The fragments

XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

XX or allelic variants are useful in the preparation of a medicament for

XX treating a disorder involving inhibiting angiogenesis in a tissue, where

XX the angiogenesis is mediated by one or more endothelial cell integrins or

XX one or more endothelial cell integrin subunits; or by promoting or

XX inducing endothelial cell apoptosis in a tissue, where the endothelial

XX cell apoptosis is mediated by one or more endothelial cell integrins or

XX one or more endothelial cell integrin subunits. The medicament is useful

XX in inhibiting tumour growth and for the regression of an established

XX tumour. The present sequence represents the amino acid sequence of human

XX type IV collagen alpha 3 chain mutant, Tumstatin 334, which consists of

XX residues 2-245 of Tumstatin. Note: The present sequence is not shown in

XX the specification but is derived from the wild type human Tumstatin

XX sequence given in figure 18A (see AAU75595)

XX

SQ Sequence 244 AA;

Query Match 100.0%; Score 244; DB 5; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.2e-241;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGSGSPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60

DB 1 GLKGRGSGSPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60

QY 61 LGTLGSLQRFMTTPELFCNVNDVCFNFRNDYSYWLSTPALMPMNPATITGRALEPYIS 120

DB 61 LGTLGSLQRFMTTPELFCNVNDVCFNFRNDYSYWLSTPALMPMNPATITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

QY 181 EFRASPFLCHGRGTCNYNSYSFSLASLNPFRMFRKPISTVKAGELEKIIISRCQVCM 240
 Db 181 EFRASPFLCHGRGTCNYNSYSFSLASLNPFRMFRKPISTVKAGELEKIIISRCQVCM 240
 QY 241 KKRH 244
 Db 241 KKRH 244

RESULT 3

ADA20225

ID ADA20225 standard; protein; 244 AA.

XX AC ADA20225;

XX AC ADA20225;

DT 20-NOV-2003 (first entry)

XX Human type IV collagen alpha 3 chain partial protein sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX metastasis; basement membrane organisation; type IV collagen network;

KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

KW cytotactic; gene therapy; alpha 3 chain; tumstatin; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..244

FT Region /note= "Tumstatin"

FT Region 1..124

FT Region /note= "Tumstatin 333; pET22b-alpha 3 (IV) NC1 region"

FT Region 1..19

FT Region /note= "T1 peptide"

FT Region 28..42

FT Region /note= "First Goodpasture epitope"

FT Region 53..72

FT Region /note= "T2 peptide"

FT Region 58..88

FT Region /note= "T3 peptide"

FT Region 83..102

FT Region /note= "T4 peptide"

FT Region 98..116

FT Region /note= "T5 peptide"

FT Region 113..132

FT Region /note= "T6 peptide"

FT Region 125..244

FT Region /note= "Tumstatin 334"

FT Region 139..152

FT Region /note= "Second Goodpasture epitope"

XX WO200305257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WFI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor

XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 52; Fig 18; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the partial amino acid sequence of the alpha 3 chain of human
 CC type IV collagen. The "tumstatin" protein of the invention was derived
 CC from this protein and comprises the full length of the present sequence.

XX Sequence 244 AA;

Query Match 100.0%; Score 244; DB 6; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.2e-241;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKKGDSGSPATWTRGTFVTRHSQTATPSCDEGTVPYSGFSLFVGNORAHQGD 60

Db 1 GLKKGDSGSPATWTRGTFVTRHSQTATPSCDEGTVPYSGFSLFVGNORAHQGD 60

QY 61 LGTLGSLQRFITMPFLFCNVNDVCFASRNDYSYWLSTPALMPNMAPITGRALEPYIS 120

Db 61 LGTLGSLQRFITMPFLFCNVNDVCFASRNDYSYWLSTPALMPNMAPITGRALEPYIS 120

QY 121 RCTVCGSPALATAVHSQTTDIPCPGHWSLWKGFSTMTSAGSEGTGALASPGSCLE 180

Db 121 RCTVCGSPALATAVHSQTTDIPCPGHWSLWKGFSTMTSAGSEGTGALASPGSCLE 180

QY 181 EFRASPFLCHGRGTCNYNSYSFSLASLNPFRMFRKPISTVKAGELEKIIISRCQVCM 240

Db 181 EFRASPFLCHGRGTCNYNSYSFSLASLNPFRMFRKPISTVKAGELEKIIISRCQVCM 240

QY 241 KKRH 244

Db 241 KKRH 244

RESULT 4

AAV67942

ID AAV67942 standard; protein; 245 AA.

XX AC AAV67942;

XX 03-APR-2000 (first entry)

XX Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.

XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;

XX benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;

XX ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;

XX myocardial angiogenesis; plaque neovascularisation; angiofibroma;

XX atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;

XX contraception; obesity.

XX Homo sapiens.

XX WO9965940-A1.

XX 23-DEC-1999.

XX 17-JUN-1999; 99WO-US013737.

XX 17-JUN-1998; 98US-0089689P.

XX 25-MAR-1999; 99US-0126175P.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2000-097708/08.
XX DR N-PSDB; AA257158.
XX PT Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or 3
XX PT chain of Type IV collagen used in, e.g. treatment of benign tumors and
XX PT rheumatoid arthritis.
XX PS Claim 32; Fig 16B; 117pp; English.
XX CC The present sequence represents the human type IV collagen alpha 3 chain.
XX CC The present invention describes an isolated protein chosen from the NC1
XX CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a
XX CC fragment, analogue, derivative or mutant, which has anti-angiogenic
XX CC properties. The anti-angiogenic proteins, multimers and chimeras are
XX CC useful for inhibiting angiogenic activity in mammalian tissue, especially
XX CC for treating diseases chosen from angiogenesis-dependent cancers, benign
XX CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular
XX CC angiogenesis diseases, Osler-Webber Syndrome, myocardial angiogenesis,
XX CC plaque neovascularisation, telangiectasia, haemophilic joints,
XX CC angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,
XX CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori
XX CC ulcers, dialysis graft vascular access stenosis, contraception and
XX CC obesity. The compositions can be used to inhibit a disease characterised
XX CC by angiogenic activity, in conjunction with radiation therapy,
XX CC chemotherapy or immunotherapy
XX SQ Sequence 245 AA;
Query Match 100.0%; Score 244; DB 3; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.2e-241;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKRGDGS PATWTRTRGFVTRHSQT TAIPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
Db 2 GLKRGDGS PATWTRTRGFVTRHSQT TAIPSCPEGTVPVLYSGFSFLVQGNQRAHQD 61
QY 61 LGTGLSCLO RFTT MPELF CNVDVNCNFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 120
Db 62 LGTGLSCLO RFTT MPELF CNVDVNCNFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 121
QY 121 RCTVCEGPAIAVHSQT DIPPCHGWISLWKGFNFMTSAGSEGTGQALASPGSCLE 180
Db 122 RCTVCEGPAIAVHSQT DIPPCHGWISLWKGFNFMTSAGSEGTGQALASPGSCLE 181
QY 181 EFRASPFLCHGRGTCNYYSNSYSFWLASLNPERMFRKPIPTVXAGELEKIISRQVCM 240
Db 182 EFRASPFLCHGRGTCNYYSNSYSFWLASLNPERMFRKPIPTVXAGELEKIISRQVCM 241
QY 241 KKH 244
Db 242 KKH 245
RESULT 5
AAU75589
ID AAU75589 standard; protein; 245 AA.
XX AC AAU75589;
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain, 'Tumstatin'.
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX KW non-Goopasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX KW Tumstatin; angiogenesis; tumour.

OS Homo sapiens.
XX FN WO200151523-A2.
XX PD 19-JUL-2001.
XX PF 08-JAN-2001; 2001WO-US0000565.
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PA Kalluri R;
XX PI WPI; 2002-189037/24.
XX DR N-PSDB; ABK15365.
XX PT A non-Goopasture fragment of alpha3(IV)NC1 domain used in detecting and
XX PT treating disorders involving angiogenesis.
XX PS Claim 29; Fig 18B; 205pp; English.
XX CC The invention relates to a non-Goopasture fragment of alpha3(IV)NC1
XX CC domain, having one or more of the characteristics selected from: (a) the
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX CC proliferation of endothelial cells; and (c) the ability to cause
XX CC apoptosis of endothelial cells. Also described are the following: (1) use
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX CC analogue or allelic variant in the preparation of a medicament for
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX CC where the angiogenesis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,
XX CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
XX CC preparation of a medicament for inhibiting angiogenesis or cell
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation
XX CC of a medicament for treating a proliferative disease in a vertebrate,
XX CC where the disease is characterised by angiogenesis that is mediated by
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and
XX CC (5) use of integrins in the preparation of a medicament for promoting or
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments
XX CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX CC or allelic variants are useful in the preparation of a medicament for
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where
XX CC the angiogenesis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits; or by promoting or
XX CC inducing endothelial cell apoptosis in a tissue, where the endothelial
XX CC cell apoptosis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits. The medicament is useful
XX CC in inhibiting tumour growth and for the regression of an established
XX CC tumour. The present sequence represents the amino acid sequence of human
XX CC type IV collagen alpha 3 chain
XX SQ Sequence 245 AA;

Query Match 100.0%; Score 244; DB 5; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.2e-241;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKRGDGS PATWTRTRGFVTRHSQT TAIPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
Db 2 GLKRGDGS PATWTRTRGFVTRHSQT TAIPSCPEGTVPVLYSGFSFLVQGNQRAHQD 61
QY 61 LGTGLSCLO RFTT MPELF CNVDVNCNFASRNDYSYWLSTPALMPMNMNAPITGRALEPYIS 120

DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.
XX Human; type IV collagen alpha 3 chain; cytosolic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutin; mutant.
XX Homo sapiens.
OS
XX WO200151523-A2.
PN 19-JUL-2001.
PD
XX 08-JAN-2001; 2001WO-US000565.
PF
XX 07-JAN-2000; 2000US-00479118.
PR
XX 04-APR-2000; 2000US-00543371.
PR
XX 21-JUL-2000; 2000US-00625191.
PR
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA
XX Kalluri R;
PI
XX WPI; 2002-188037/24.
DR
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
PT treating disorders involving angiogenesis.
PT
XX Example 32; Page; 205pp; English.
PS
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha4, alpha5, alpha6, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of
CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in
CC the specification but is derived from the wild type human Tumstatin
CC sequence given in figure 18A (see AAU75589)
XX Sequence 191 AA;

Query Match 78.3%; Score 191; DB 5; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.3e-187;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 NQRAHGDGLTGSLCQRFTTTPFLFCNVNDVNCNPNASNDYSYWLSTPALMPNMAPITG 112
DB 1 NQRAHGDGLTGSLCQRFTTTPFLFCNVNDVNCNPNASNDYSYWLSTPALMPNMAPITG 60
QY 113 RALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQAL 172
DB 61 RALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQAL 120
QY 173 ASGSCLEBPRAFPFLECHRGTCNVYNSYSYFWLASLNPERMERKPISTVKAGLEKI 232
DB 121 ASGSCLEBPRAFPFLECHRGTCNVYNSYSYFWLASLNPERMERKPISTVKAGLEKI 180
QY 233 ISRCQVCMKKR 243
DB 181 ISRCQVCMKKR 191
RESULT 10
ADA20260
ID ADA20260 standard; protein; 191 AA.
XX
AC ADA20260;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human tumstatin deletion protein tum-1 amino acid sequence.
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;
KW tumstatin N53.
XX
OS Homo sapiens.
XX
PN WO2003059257-A2.
XX
PD 24-JUL-2003.
XX
PF 20-DEC-2002; 2002WO-US040938.
XX
PR 21-DEC-2001; 2001US-00032221.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
DR WPI; 2003-587256/55.
DR N-PSDB; ADA20224.
XX
PT New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Claim 94; SEQ ID NO 22; 240pp; English.
XX
CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumor growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC

CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tum-1 (tumstatin N53), an abridged form of the
CC "tumstatin" protein of the invention which was derived from the amino
CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This
CC sequence (Seq ID22) does not appear in the specification but was created
CC by the indexer from information given in the specification.
XX
SQ

Sequence 191 AA;

Query Match 78.3%; Score 191; DB 6; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.3e-187;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 54 QRAHQDGLTGLSCQRFMTMFLFCNVNDVNCNFAASNDYSVWLSPTALPMNVPATIGR 113
DB 1 QRAHQDGLTGLSCQRFMTMFLFCNVNDVNCNFAASNDYSVWLSPTALPMNVPATIGR 60
QY 114 ALEPVISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFMFTSAGSEGTGQALA 173
DB 61 ALEPVISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFMFTSAGSEGTGQALA 120
QY 174 SPGSCLEBRASPFLFCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTSTVKAGELEKII 233
DB 121 SPGSCLEBRASPFLFCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTSTVKAGELEKII 180
QY 234 SRCQVCMKKRH 244
DB 181 SRCQVCMKKRH 191

RESULT 11
AAU75598
ID AAU75598 standard; protein; 254 AA.
AC AAU75598;
XX

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tum-3.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX

OS Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

PT treating disorders involving angiogenesis.

XX Example 36; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tum-3, which consists of residues
CC 133-244 of Tumstatin. Note: The present sequence is not shown in the
CC specification but is derived from the wild type human Tumstatin sequence
CC given in figure 18A (see AAU75598)
XX

SQ Sequence 254 AA;

Query Match 75.0%; Score 183; DB 5; Length 254;
Best Local Similarity 100.0%; Pred. No. 4.7e-179;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LGTGLSCQRFMTMFLFCNVNDVNCNFAASNDYSVWLSPTALPMNVPATIGRALEPYIS 120

DB 72 LGTGLSCQRFMTMFLFCNVNDVNCNFAASNDYSVWLSPTALPMNVPATIGRALEPYIS 131

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFMFTSAGSEGTGQALASPGSCLE 180

DB 132 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFMFTSAGSEGTGQALASPGSCLE 191

QY 181 EFRASPFLECHGRGTCNYSNSYSFWLASLNPERMFRKPIPTSTVKAGELEKIIISRCQVCM 240

DB 192 EFRASPFLECHGRGTCNYSNSYSFWLASLNPERMFRKPIPTSTVKAGELEKIIISRCQVCM 251

QY 241 KKR 243

DB 252 KKR 254

RESULT 12

AAV31993

ID AAV31993 standard; protein; 268 AA.

XX

AC AAV31993;

XX 05-JAN-2000 (first entry)

XX Type IV collagen NC1 domain alpha-3 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;

KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;

KW rheumatoid arthritis; retinal neovascularization;
 KW choroidal neovascularization; macular degeneration;
 KW corneal neovascularization; retinopathy of prematurity;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidemic keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; soggren's, acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW ulcer; herpes simplex infection; Herpes zoster infection;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;
 KW sarcoed; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Bales disease; Bechets disease;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu;
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
 KW pemphigoid.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..17
 FT /note= "BM40 signal peptide"
 FT Protein 18..268
 FT /note= "mature protein"
 FT Peptide 18..25
 FT /note= "affinity tag"
 FT Protein 26..268
 FT /note= "NC1 alpha-3 monomer"
 XX
 XX WO9949885-A2.
 XX
 XX 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WO-US006445.
 XX
 XX 27-MAR-1998; 98US-0079783P.
 XX 29-OCT-1998; 98US-0106170P.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 XX Hudson BG, Sarraz MP;
 XX
 XX WPI; 1999-601297/51.
 XX N-PSDB; AA220091.
 XX
 XX Inhibition of angiogenesis with non-collagenous alpha chain monomer
 XX useful for treating e.g. tumor growth or metastasis, neovascularisation,
 XX etc.
 XX
 XX Disclosure; Fig 17c; 56pp; English.
 XX
 XX This sequence represents a recombinant type IV collagen non-collagenous
 XX (NC1) domain alpha-3 polypeptide composed of a BM40 signal sequence
 XX (which is cleaved from the mature protein) to facilitate protein
 XX secretion, and a mature protein comprising an affinity tag (facilitates
 XX purification and identification of the material) and the alpha-1 chain
 XX monomer. The invention provides methods and kits for inhibiting
 XX angiogenesis, tumor growth and metastasis, and endothelial cell
 XX interaction with the extracellular matrix, each method comprising
 XX contacting the tumor or animal tissue with 1 or more isolated type IV
 XX collagen NC1 alpha chain monomer(s) selected from the group consisting of
 XX alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-
 XX 96). The monomers can be produced via recombinant protein expression. The
 XX polynucleotides and polypeptides are used to treat an angiogenesis-
 XX mediated disorder or condition, especially selected from solid and blood-
 XX borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal

CC neovascularization, choroidal neovascularization, macular degeneration,
 CC corneal neovascularization, retinopathy of prematurity, corneal graft
 CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
 CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, soggren's,
 CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
 CC simplex infections, herpes zoster infections, protozoan infections,
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
 CC keratolysis, scleritis, systemic lupus, polyarteritis, Wegener's
 CC sarcoidosis, sickle cell anaemia, sarcoed, pseudoxanthoma elasticum, radial keratotomy,
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
 CC disease, myopia, optic pits, Stargarts disease, Bales disease, Bechets
 CC disease, vitritis, Lyme's disease, disease, Eales disease, chronic
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
 CC complications, abnormal proliferation of fibrovascular tissue,
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)
 XX
 SQ Sequence 268 AA;
 Query Match 66.8%; Score 163; DB 2; Length 268;
 Best Local Similarity 100.0%; Pred. No. 1.6e-158;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 KRGDGSPATWTRGFVTRHSQTTPSCPEGTVELYSGFSLFVQGNQRAHQDGLTL 64
 Db 29 KRGDGSPATWTRGFVTRHSQTTPSCPEGTVELYSGFSLFVQGNQRAHQDGLTL 88
 Qy 65 GSCLOQRTTTPFLFCNVNDVCFNSNDYSYMLSTPALMPMNPAPITGRALPEYISRCTV 124
 Db 89 GSCLOQRTTTPFLFCNVNDVCFNSNDYSYMLSTPALMPMNPAPITGRALPEYISRCTV 148
 Qy 125 CEGPAIAIAVHSQTTPDIPPCPHGWISLWKGFSPIMFTSAGSEG 167
 Db 149 CEGPAIAIAVHSQTTPDIPPCPHGWISLWKGFSPIMFTSAGSEG 191
 RESULT 13
 AAY97555
 ID AAY97555 standard; protein; 268 AA.
 XX
 AC AAY97555;
 XX
 DT 12-FEB-2001 (first entry)
 XX
 XX Human alpha3(IV)NC1 protein sequence.
 DE
 XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
 KW tumor growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoed;
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
 KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1.
 XX
 OS Homo sapiens.
 XX
 XX WO200059532-A1.
 XX
 XX 12-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US008678.
 XX
 XX 01-APR-1999; 99US-0127391P.
 XX
 XX (BIOS-) BIOSTRATUM INC.
 XX
 XX Brooks P, Hudson B;
 XX

```

DR WPI; 2000-664962/64.
DR N-PSDB; AAA90593.
XX
PT Use of antagonists of specific integrin receptors for inhibiting
PT angiogenesis, tumor growth or metastases, or endothelial cell
PT interactions with the extracellular matrix.
XX
XX
PS Disclosure; Fig 17c; 78pp; English.
XX
CC This sequence is a human type IV collagen alpha chain monomer, designated
CC alpha3(I)NC1. The invention relates to a method for inhibiting
CC angiogenesis, tumor growth or metastases, or endothelial cell
CC interactions with the extracellular matrix, comprising contacting the
CC cells or tissue with a polypeptide composition containing antagonists of
CC specific integrin receptors. The methods and the antagonists are useful
CC for inhibiting angiogenesis, tumor growth or metastases, or endothelial
CC cell interaction with the extracellular matrix. The antagonists are also
CC useful for treating diseases and conditions with accompanying undesired
CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
CC treating non-tumorigenic diseases and conditions with accompanying
CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
CC retinal neovascularisation, choroidal neovascularisation, muscular
CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
CC complications, chronic inflammation or psoriasis
XX
SQ Sequence 268 AA;
Query Match 66.8%; Score 163; DB 3; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.6e-158;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 KRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQDGLTL 64
Db 29 KRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQDGLTL 88
QY 65 GSCLQRFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPNMAPITGRALEPIVSRCTV 124
Db 89 GSCLQRFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPNMAPITGRALEPIVSRCTV 148
QY 125 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFNFMTSAGSEG 167
Db 149 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFNFMTSAGSEG 191
RESULT 14
AA95918
ID AA95918 standard; protein; 211 AA.
XX
XX AA95918;
XX
XX 20-NOV-2000 (first entry)
XX
XX Human Goodpasture antigen Deltav.
XX
XX Goodpasture antigen; GPdeltav; goodpasture antigen binding protein; GPBP;
XX human; autoimmune disease; apoptosis; cancer; tumour; therapy.
XX
XX Homo sapiens.
XX
XX WO2000050607-A2.
XX
XX 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-IB000324.
XX
XX 24-FEB-1999; 99US-0121483P.
XX
XX (SAUS/) SAUS J.
XX

```

```

PI Saus J;
XX
DR WPI; 2000-572094/53.
DR N-PSDB; AAA50367.
XX
PT Novel Goodpasture antigen binding proteins useful for diagnosing and
PT treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX
XX
PS Claim 36; Page 151-152; 159pp; English.
XX
CC The present sequence is that of human recombinant Goodpasture antigen
CC (GP) Deltav, i.e. an alternative form of human GP resulting from splicing
CC out of exon V. The recombinant protein, lacking the Met-1 residue, was
CC expressed in bacterial pellets using modified vector pET15b carrying
CC GPdeltav cDNA (see AAA50367). The invention relates to novel Goodpasture
CC antigen binding proteins (GPBPs, see AAY95900-11), which bind to and
CC phosphorylate the unique N-terminal region of human GP, and which are
CC highly expressed in several autoimmune conditions. Claimed methods for
CC treating an autoimmune disorder, cell apoptosis or a tumour involve
CC modifying the expression or activity of GPBP, especially using a GP-
CC derived peptide, such as GPdeltav
XX
SQ Sequence 211 AA;
Query Match 65.2%; Score 159; DB 3; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.7e-154;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKKGSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60
Db 1 GLKKGSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60
QY 61 LGTLGSLQRFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPNMAPITGRALEPYIS 120
Db 61 LGTLGSLQRFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPNMAPITGRALEPYIS 120
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFNFMTSAGSEG 159
Db 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFNFMTSAGSEG 159
RESULT 15
ABG79208
ID ABG79208 standard; protein; 211 AA.
XX
XX ABG79208;
XX
XX 15-NOV-2002 (first entry)
XX
XX Human GP protein isoform GPdeltav.
XX
XX Goodpasture antigen binding protein; Goodpasture syndrome;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX
XX Homo sapiens.
XX
XX WO200261430-A2.
XX
XX 08-AUG-2002.
XX
XX 31-JAN-2002; 2002WO-EP001010.
XX
XX 31-JAN-2001; 2001US-0265249P.
XX
XX (SAUS/) SAUS J.
XX
XX Saus J;
XX
XX WPI; 2002-619280/66.
XX

```

```
DR N-PSDB; ABS64491.
XX
XX Identifying candidate compounds for treating autoimmune conditions, e.g.
PT Goodpasture syndrome or lupus, comprises identifying compounds that
PT reduce phosphorylation of, or formation of conformational isomers of,
PT target proteins.
XX
XX Example 3; Page 199-200; 217pp; English.
XX
XX The invention relates to identifying candidate compounds to treat an
XX autoimmune condition by identifying compounds that reduce phosphorylation
XX of a first target protein (I) which is selected from Goodpasture antigen
XX binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCL)
XX domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-
XX Ala-Thr-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
XX Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
XX conformational isomers of the second target protein (II) (selected from
XX an alpha3 type IV collagen NCL domain polypeptide and myelin basic
XX protein, MBP). Also included are (1) an isolated type IV collagen alpha3
XX NCL domain conformational isomer, which has an amino acid sequence
XX identical to the wild type alpha3 type IV collagen NCL domain, is
XX stabilised by disulphide bonds, and has a molecular weight in a non-
XX reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
XX type IV collagen alpha3 NCL domain. The human gene for GPBP is located on
XX chromosome 5q13. The method is useful for treating autoimmune conditions,
XX such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
XX lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
XX sequence represents an alpha3 type IV collagen non-collagenous (NCL)
XX domain (also known as the GP antigen) or an MBP isoform
XX
XX Sequence 211 AA;
XX
XX Query Match 65.2%; Score 159; DB 5; Length 211;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-154;
XX Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GLKGRGSGSPATWTRGTFVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQD 60
XX DB 1 GLKGRGSGSPATWTRGTFVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQD 60
XX
XX QY 61 LGTLGSLQRFMTTFPLFCNVNDVNCNFRNDYSYWLSTPALMPMNPATITGRALEPIYS 120
XX DB 61 LGTLGSLQRFMTTFPLFCNVNDVNCNFRNDYSYWLSTPALMPMNPATITGRALEPIYS 120
XX
XX QY 121 RCTVCEGPAIAVAHSQTIDPPCPHGWSLWKGFSPFM 159
XX DB 121 RCTVCEGPAIAVAHSQTIDPPCPHGWSLWKGFSPFM 159
XX
XX RESULT 16
XX ADCL7697
XX ID ADCL7697 standard; protein; 232 AA.
XX
XX AC ADCL7697;
XX
XX XT 18-DBS-2003 (first entry)
XX
XX DE Human type IV collagen alpha 3 chain protein SEQ ID NO:304.
XX
XX KW crystallised NCL domain hexamer of type IV collagen;
XX KW angiogenesis inhibitor; angiogenesis-mediated disease;
XX KW tumour metastasis inhibitor; tumour growth inhibitor;
XX KW endothelial cell interaction inhibitor;
XX KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
XX KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX KW endothelial cell adhesion inhibitor;
XX KW endothelial cell proliferation inhibitor;
XX KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX KW blood-borne tumour.
XX
XX OS Homo sapiens.
XX
```

```
PN WO2003012122-A2.
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
XX
XX 29-OCT-2001; 2001US-0351289P.
XX
XX 22-MAR-2002; 2002US-0366854P.
XX
XX 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Disclosure; SEQ ID NO 304; 168pp; English.
XX
XX The present invention describes a crystallised NCL domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (6) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCL
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCL domain hexamer of type IV collagen (I) has cytostatic,
XX antipsoriatic, antianemic, ophthalmological, antiarteriosclerotic and
XX antiulcer activities, and can be used as an inhibitor of angiogenesis in
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents an amino acid sequence which is used in the exemplification of
XX the present invention.
XX
XX SQ Sequence 232 AA;
XX
XX Query Match 63.5%; Score 155; DB 7; Length 232;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-150;
XX Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 13 ATWTRGTFVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQDGLTGLSCLQRF 72
XX DB 1 ATWTRGTFVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQDGLTGLSCLQRF 60
XX
XX QY 73 TWPFLECNVNDVNCNFRNDYSYWLSTPALMPMNPATITGRALEPIYSRCTVCEGPAIAI 132
XX DB 61 TWPFLECNVNDVNCNFRNDYSYWLSTPALMPMNPATITGRALEPIYSRCTVCEGPAIAI 120
XX
XX QY 133 AVHSQTIDPPCPHGWSLWKGFSPFMFTSAGSEG 167
XX DB 121 AVHSQTIDPPCPHGWSLWKGFSPFMFTSAGSEG 155
XX
XX RESULT 17
XX AAY44172
XX ID AAY44172 standard; protein; 218 AA.
```

```
XX AC AAY44172;
XX DT 01-FEB-2000 (first entry)
XX DE Human type IV collagen alpha3 chain protein.
XX KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;
XX KW Goodpasture syndrome; antibody; Blood; tissue; human; nephrotrophism.
XX OS Homo sapiens.
XX PN US5973120-A.
XX PD 26-OCT-1999.
XX PF 07-MAR-1995; 95US-00399889.
XX PR 30-NOV-1990; 90US-00621091.
XX PA (UYVA ) UNIV YALE.
XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX PI Hudson BG, Reiders ST, Morrison KE;
XX DR WPI; 1999-610317/52.
XX DR N-PSDB; AAZ28775.
XX PT Isolated alpha 3 chain of type IV collagen polypeptide useful for
XX PT diagnosis and treatment of Goodpasture syndrome.
XX PS Claim 2; Col 35-36; 27pp; English.
XX CC This sequence represents a recombinant human alpha3 chain of type IV
XX CC collagen polypeptide. The sequence corresponds to the 218 amino acids of
XX CC the C-terminal non-collagenous domain. Alpha3 chain collagen polypeptides
XX CC are useful for detecting Goodpasture antibodies in blood or tissue from a
XX CC human patient and for treating Goodpasture syndrome, especially by
XX CC neutralising the antibodies in the blood. The polypeptides also have a
XX CC nephrotrophic activity
XX SQ Sequence 218 AA;
Query Match 57.8%; Score 141; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 QTTAIPSCPEGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLOQFTTTPFLFCNVNDVCN 85
Db 1 QTTAIPSCPEGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLOQFTTTPFLFCNVNDVCN 60
QY 87 FASRNDYSYWLSTPALMPMNPAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
Db 61 FASRNDYSYWLSTPALMPMNPAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120
QY 147 GWISLWKGFSPIMFTSAGSEG 167
Db 121 GWISLWKGFSPIMFTSAGSEG 141
RESULT 18
AAY56784
ID AAY56784 standard; protein; 218 AA.
XX AC AAY56784;
XX DT 27-MAR-2000 (first entry)
XX DE Human alpha3 type IV collagen C-terminal domain.
XX KW Goodpasture syndrome; type IV collagen; alpha3 chain; human.
XX OS Homo sapiens.
```

```
XX PN US6007980-A.
XX PD 28-DEC-1999.
XX PF 07-OCT-1998; 98US-00167364.
XX PR 30-NOV-1990; 90US-00621091.
XX PR 07-MAR-1995; 95US-00399889.
XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX PA (UYVA ) UNIV YALE.
XX PI Hudson BG, Reiders ST, Morrison KE;
XX DR WPI; 2000-096371/08.
XX DR N-PSDB; AAZ46729.
XX PT Diagnosing and treating Goodpasture syndrome using a peptide derived from
XX PT type IV collagen.
XX PS Disclosure; Col 23-26; 26pp; English.
XX CC The invention provides a method of detecting Goodpasture antibodies in
XX CC the fluid of a patient by contacting it with a peptide comprising at most
XX CC 218 amino acids of the human alpha3 chain type IV collagen that contains
XX CC the fragment shown in AAY56785. The methods are useful for the diagnosis
XX CC and treatment of Goodpasture syndrome. The present sequence represents
XX CC the carboxy terminal noncollagenous domain of the human alpha3 chain of
XX CC type IV collagen
XX SQ Sequence 218 AA;
Query Match 57.8%; Score 141; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 QTTAIPSCPEGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLOQFTTTPFLFCNVNDVCN 86
Db 1 QTTAIPSCPEGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLOQFTTTPFLFCNVNDVCN 60
QY 87 FASRNDYSYWLSTPALMPMNPAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
Db 61 FASRNDYSYWLSTPALMPMNPAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120
QY 147 GWISLWKGFSPIMFTSAGSEG 167
Db 121 GWISLWKGFSPIMFTSAGSEG 141
RESULT 19
AAE09484
ID AAE09484 standard; protein; 218 AA.
XX AC AAE09484;
XX DT 19-NOV-2001 (first entry)
XX DE Human alpha-3 chain of type IV collagen protein.
XX KW Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
XX KW Goodpasture syndrome.
XX OS Homo sapiens.
XX PN US6277558-B1.
XX PD 21-AUG-2001.
XX PF 12-NOV-1999; 99US-00439897.
XX PR 30-NOV-1990; 90US-00621091.
XX PR 07-MAR-1995; 95US-00399889.
```

```
PR 07-OCT-1998; 98US-00167364.
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX Hudson BG;
XX WPI; 2001-540401/60.
DR N-PSDB; RAD16400.
XX
PT Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
PT Goodpasture antibodies from bodily fluid/tissue from patient or for
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with
PT the polypeptide.
XX
XX Example 4; Col 37-40; 46pp; English.
XX
CC The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is human alpha-3 chain of type IV collagen protein
XX
XX Sequence 218 AA;
Query Match 57.8%; Score 141; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 QTAIPSCPEGTVPVLYSGFSLFVQGNQRAHQDGLTGLSCLOQRTTMTPLFCNVNDVCN 86
DB 1 QTAIPSCPEGTVPVLYSGFSLFVQGNQRAHQDGLTGLSCLOQRTTMTPLFCNVNDVCN 60
QY 87 FASRNDYSYWLSTPALMPMNMAPITGRALEFYISRCTVCEGPAIAIAVHSQTTDIPCPH 146
DB 61 FASRNDYSYWLSTPALMPMNMAPITGRALEFYISRCTVCEGPAIAIAVHSQTTDIPCPH 120
QY 147 GWISLWKGFSPIMFISAGSEG 167
DB 121 GWISLWKGFSPIMFISAGSEG 141
RESULT 20
ADA20261
ID ADA20261 standard; protein; 132 AA.
XX
XX ADA20261;
AC
XX
XX 20-NOV-2003 (first entry)
XX Human tumstatin deletion protein tum-2 amino acid sequence.
XX
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.
XX
XX Homo sapiens.
XX WO2003059257-A2.
XX
XX 24-JUL-2003.
XX
XX 20-DEC-2002; 2002WO-US040938.
XX
XX 21-DEC-2001; 2001US-00032221.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX
```

```
XX WPI; 2003-587256/55.
DR N-PSDB; ADA20224.
XX
XX New peptide, useful for preparing a composition for inhibiting tumour
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
XX Claim 94; SEQ ID NO 23; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tum-2, an abridged form of the "tumstatin" protein of
XX 3 chain of human type IV collagen. Note: This sequence (Seq ID33) does
XX not appear in the specification but was created by the indexer from
XX information given in the specification.
XX
XX Sequence 132 AA;
Query Match 54.1%; Score 132; DB 6; Length 132;
Best Local Similarity 100.0%; Pred. No. 5.6e-127;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKXRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60
DB 1 GLKXRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSCLOQRTTMTPLFCNVNDVCNDFASRNDYSYWLSTPALMPMNMAPITGRALEPYIS 120
DB 61 LGTLGSCLOQRTTMTPLFCNVNDVCNDFASRNDYSYWLSTPALMPMNMAPITGRALEPYIS 120
QY 121 RCTVCEGPAIAI 132
DB 121 RCTVCEGPAIAI 132
RESULT 21
AAU75597
ID AAU75597 standard; protein; 132 AA.
XX
XX AAU75597;
XX
XX 08-MAY-2002 (first entry)
XX Human type IV collagen alpha 3 chain mutant, Tum-2.
XX
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX non-goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canastatin;
XX Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
XX Homo sapiens.
XX WO200151523-A2.
XX
XX 19-JUL-2001.
XX
XX 08-JAN-2001; 2001WO-US0000565.
XX
```


PR 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX Kalluri R;
 XX WPI; 2002-188037/24.
 DR A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.
 XX Claim 31; Page 152; 205pp; English.
 XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the
 CC specification but is derived from the wild type human Tumstatin sequence
 CC given in figure 10A (see AAU75599)
 XX Sequence 132 AA;
 SQ
 Query Match 53.7%; Score 131; DB 5; Length 132;
 Best Local Similarity 100.0%; Pred. No. 5.9e-126;
 Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKKGKDGSPATWTRGTVFTRHSQTTAIPSCPEGTVPLVSGSFVFGVQNGAHOQD 60
 Db 2 GLKKGKDGSPATWTRGTVFTRHSQTTAIPSCPEGTVPLVSGSFVFGVQNGAHOQD 61
 QY 61 LGTLGSLQRFMTMPFLFNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALPEYIS 120
 Db 62 LGTLGSLQRFMTMPFLFNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALPEYIS 121
 QY 121 RCTVCBGPAA 131
 Db 122 RCTVCBGPAA 132

RESULT 22
 AAU75594
 ID AAU75594 standard; protein; 124 AA.
 XX AC AAU75594;
 XX DT 08-MAY-2002 (first entry)
 XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.
 XX KW Human; type IV collagen alpha 3 chain; cystostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten, Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutant;
 XX OS Homo sapiens.
 XX PN WO200151523-A2.
 XX PD 19-JUL-2001.
 XX PF 08-JAN-2001; 2001WO-US0000565.
 XX PR 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 Kalluri R;
 WPI; 2002-188037/24.
 A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 treating disorders involving angiogenesis.
 Example 33; Page: 205pp; English.
 The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 domain, having one or more of the characteristics selected from: (a) the
 ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 proliferation of endothelial cells; and (c) the ability to cause
 apoptosis of endothelial cells. Also described are the following: (1) use
 of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 analogue or allelic variant in the preparation of a medicament for
 treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 where the angiogenesis is mediated by one or more endothelial cell
 integrins or one or more endothelial cell integrin subunits; or (b) by
 promoting or inducing endothelial cell apoptosis in a tissue, where the
 endothelial cell apoptosis is mediated by one or more endothelial cell
 integrins or one or more endothelial cell integrin subunits; (2) use of
 an antibody or peptide that specifically binds the alpha1, alpha2,
 alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 preparation of a medicament for inhibiting angiogenesis or cell
 proliferation; (3) use of an inhibitor, such as an antibody, antibody
 fragment or peptide of receptor-mediated angiogenesis in the preparation
 of a medicament for treating a proliferative disease in a vertebrate,
 where the disease is characterised by angiogenesis that is mediated by
 receptors to Arresten, Canstatin or Tumstatin and where the receptors
 inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 the presence of a medicament for promoting angiogenesis in a tissue; and
 (5) use of integrins in the preparation of a medicament for promoting or
 inducing angiogenesis or cell proliferation in a tissue. The fragments
 of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 or allelic variants are useful in the preparation of a medicament for
 treating a disorder involving inhibiting angiogenesis in a tissue, where
 the angiogenesis is mediated by one or more endothelial cell integrins or
 one or more endothelial cell integrin subunits; or by promoting or
 inducing endothelial cell apoptosis in a tissue, where the endothelial
 cell apoptosis is mediated by one or more endothelial cell integrins or
 one or more endothelial cell integrin subunits. The medicament is useful
 in inhibiting tumour growth and for the regression of an established
 tumour. The present sequence represents the amino acid sequence of human
 type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues
 1-132 of Tumstatin. Note: The present sequence is not shown in the
 specification but is derived from the wild type human Tumstatin sequence
 given in figure 10A (see AAU75599)

CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of
CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in
CC the specification but is derived from the wild type human Tumstatin
CC sequence given in figure 18A (see AAU75589)

XX
SQ Sequence 124 AA;

Query Match 50.8%; Score 124; DB 5; Length 124;
Best Local Similarity 100.0%; Pred. No. 8.5e-119;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Dy |||||
1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120
Dy |||||
61 LGTLGSLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120
QY 121 RCTV 124
Dy |||||
121 RCTV 124

RESULT 23.

ADA20258.
ID ADA20258 standard; protein; 124 AA.

AC ADA20258;

XX 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytoskeletal; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

OS Homo sapiens.

PN WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 20; 240pp; English.

CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
CC ID20) does not appear in the specification but was created by the indexer
CC from information given in the specification.

XX SQ Sequence 124 AA;

Query Match 50.8%; Score 124; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 8.5e-119;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Dy |||||
1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120
Dy |||||
61 LGTLGSLQRFRTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120
QY 121 RCTV 124
Dy |||||
121 RCTV 124

RESULT 24

ADA20259
ID ADA20259 standard; protein; 120 AA.

XX ADA20259;

XX 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tumstatin 334 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytoskeletal; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 334.

OS Homo sapiens.

PN WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 21; 240pp; English.

CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tumstatin 334, an abridged form of the "tumstatin"
 CC protein of the invention which was derived from the amino acid sequence
 CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
 CC ID21) does not appear in the specification but was created by the indexer
 CC from information given in the specification.

XX Sequence 120 AA;

Query Match 49.2%; Score 120; DB 6; Length 120;
 Best Local Similarity 100.0%; Pred. No. 1.1e-114;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 125 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 184
 Db 1 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 60
 QY 185 SPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIPTVTKAGELEKIIISRCQVCMKKRH 244
 Db 61 SPFLECHGRGTCNYNSYSYFWLASLNPERMFRKPIPTVTKAGELEKIIISRCQVCMKKRH 120

RESULT 25

ID ADA20262 standard; protein; 112 AA.

XX ADA20262;

AC ADA20262;

DT 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tum-3 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-3.

OS Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

PF 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PI Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

PS Example 35; SEQ ID NO 24; 240pp; English.

XX

CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of tum-3, an abridged form of the "tumstatin" protein of
 CC the invention which was derived from the amino acid sequence of the alpha
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID24) does
 CC not appear in the specification but was created by the indexer from
 CC information given in the specification.

XX Sequence 112 AA;

Query Match 45.9%; Score 112; DB 6; Length 112;
 Best Local Similarity 100.0%; Pred. No. 1.6e-106;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 192
 Db 1 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 60

QY 193 RGTCTNYNSYSYFWLASLNPERMFRKPIPTVTKAGELEKIIISRCQVCMKKRH 244

Db 61 RGTCTNYNSYSYFWLASLNPERMFRKPIPTVTKAGELEKIIISRCQVCMKKRH 112

RESULT 26

AAAR79164

ID AAR79164 standard; protein; 218 AA.

XX AAR79164;

DT 22-DEC-1995 (first entry)

DE Partial sequence of human alpha 3 chain of type IV collagen.

XX Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.

OS Homo sapiens.

XX US5424408-A.

XX 13-JUN-1995.

PF 30-NOV-1990; 90US-00621091.

XX 30-NOV-1990; 90US-00621091.

PA (UYVA) UNIV YALE.

PA (UNIV) UNIV KANSAS MEDICAL CENT.

PI Morrison KE, Reeders ST, Hudson BG;

XX WPI; 1995-262631/34.

DR N-PSDB; AAQ96291.

PT cDNA's encoding human or bovine alpha-3 type 4 collagen peptide(s) -

PT useful for detection and therapeutic removal of antibodies associated

PS with Goodpasture syndrome.

XX Disclosure; Col 7-10; 33pp; English.

CC Using the PCR with primers derived from each end of the known 27 AA
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp bovine
 CC genomic fragment was amplified. This fragment was then used to a bovine
 CC lens cDNA library and a 1.5 kb partial cDNA clone was obtained. (clone
 CC KWC15). This encodes 238 residues of the triple helical collagenous
 CC domain and all 233 residues of the C-terminal non-collagenous (NC1)
 CC domain of the alpha 3 (IV) chain. This bovine cDNA clone was used to
 CC screen a human kidney cDNA library and a 2.7 kb human cDNA clone (clone
 CC KWC27) was obtained. This clone encodes 218 residues of the NC1 domain and a
 CC portion of the 3' UTR region of the human alpha 3 (IV) chain. The COL4A3
 CC gene localises to chromosome 2 and therefore mutations in COL4A3 cannot
 CC be responsible for Alport syndrome which is X-linked. An isolated and
 CC substantially pure nt. having the sequence in AAQ96291 is claimed
 XX
 XX Sequence 218 AA;

Query Match 40.6%; Score 99; DB 2; Length 218;
 Best Local Similarity 100.0%; Pred. No. 6.4e-93;
 Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 QTTAIPSCPGTGLYSGFGLFVQGNQRAHGQDLGLGSLQRFMTTTPFLFCNVNDVCN 86
 Db 1 QTTAIPSCPGTGLYSGFGLFVQGNQRAHGQDLGLGSLQRFMTTTPFLFCNVNDVCN 60
 QY 87 FASNDYSYMLSTPALMPMNPITGTALEPYISRCTVC 125
 Db 61 FASNDYSYMLSTPALMPMNPITGTALEPYISRCTVC 99

RESULT 27
 AAU75607
 ID AAU75607 standard; protein; 88 AA.

XX AAU75607;
 XX
 DT 08-MAY-2002 (first entry)
 XX Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.
 XX
 XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; muten; mutant.
 XX
 OS Homo sapiens.
 XX WO20015123-A2.
 XX
 PD 19-JUL-2001.

XX 08-JAN-2001; 2001WO-0000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH-ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.

XX Claim 32; Page 152; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists
 CC of residues 45-132 of Tumstatin
 XX
 XX Sequence 88 AA;

Query Match 36.1%; Score 88; DB 5; Length 88;
 Best Local Similarity 100.0%; Pred. No. 5.4e-82;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GFSFLFVQGNQRAHGQDLGLGSLQRFMTTTPFLFCNVNDVCNPFASNDYSYMLSTPALM 103
 Db 1 GFSFLFVQGNQRAHGQDLGLGSLQRFMTTTPFLFCNVNDVCNPFASNDYSYMLSTPALM 60
 QY 104 PMNMAPITGRALEPYISRCTVCEGPAIA 131
 Db 61 PMNMAPITGRALEPYISRCTVCEGPAIA 88

RESULT 28

ADA20271
 ID ADA20271 standard; protein; 88 AA.

XX ADA20271;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human;
 KW tumstatin 45-132.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

PR 21-DEC-2001; 2001US-00032221.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2003-587256/55.
DR N-PSDB; ADA2024.
XX
PT New peptide, useful for preparing a composition for inhibiting tumor
FT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Claim 94; SEQ ID NO 33; 240pp; English.
XX
CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
CC ID33) does not appear in the specification but was created by the indexer
CC from information given in the specification.
XX
SQ Sequence 88 AA;
Query Match 36.1%; Score 88; DB 6; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.4e-82;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 FSLFVQGNORAHGQDLGTGLGSLQRTFTMPFLFCNVNDVNCNPASNDYSYMLSTPALMP 104.
DB 1 FSLFVQGNORAHGQDLGTGLGSLQRTFTMPFLFCNVNDVNCNPASNDYSYMLSTPALMP 60
QY 105 MNVAPITGRALEPIYSRCTVCEGPATAI 132
DB 61 MNVAPITGRALEPIYSRCTVCEGPATAI 88
RESULT 29
AAU75608
ID AAU75608 standard; protein; 88 AA.
AC AAU75608;
XX
DT 08-MAY-2002 (first entry)
XX Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 82
FT /note= "Wild type Cys substituted with Ala"
XX
PN WO2001:51523-A2.

XX 19-JUL-2001.
PD
XX
PF 08-JAN-2001; 2001WO-US000565.
XX
PR 07-JAN-2000; 2000US-00479118.
PR 04-APR-2000; 2000US-00543371.
PR 21-JUL-2000; 2000US-00625191.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA Kalluri R;
XX
PI WPI; 2002-188037/24.
XX
DR
XX
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
PT treating disorders involving angiogenesis.
XX
PS Claim 41; Page 153; 205pp; English.
XX
CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
CC consists of residues 5-126 of Tumstatin
XX
SQ Sequence 88 AA;

Query Match 33.2%; Score 81; DB 5; Length 88;
Best Local Similarity 100.0%; Pred. No. 8.3e-75;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 44 GFSFLFVQGNORAHGQDLGTGLGSLQRTFTMPFLFCNVNDVNCNPASNDYSYMLSTPALM 103
DB 1 GFSFLFVQGNORAHGQDLGTGLGSLQRTFTMPFLFCNVNDVNCNPASNDYSYMLSTPALM 60
QY 104 PMNMAPITGRALEPIYSRCTV 124
DB 61 PMNMAPITGRALEPIYSRCTV 81

```
RESULT 30
ADA20272 ID ADA20272 standard; protein; 88 AA.
XX AC ADA20272;
XX DE 20-NOV-2003 (first entry)
XX DT Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.
XX DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX KW metastasis; basement membrane organisation; type IV collagen network;
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX KW cytoskeletal; gene therapy; alpha 3 chain; tumstatin; human;
XX KW tumstatin 5-125-C-A; mutant; mutein.
XX OS Synthetic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 81
FT FT /note= "Wild-type Cys substituted by Ala at position 125
FT FT of full-length tumstatin"
XX KW WO2003059257-A2.
XX PN 24-JUL-2003.
XX PD
XX PF 20-DEC-2002; 2002WO-US040938.
XX PR 21-DEC-2001; 2001US-00032221.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX PI WPI; 2003-587256/55.
XX DR
XX PT New peptide, useful for preparing a composition for inhibiting tumor
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX PS Claim 94; SEQ ID NO 34; 240pp; English.
XX CC This invention relates to novel isolated proteins and their fragments
XX CC with anti-angiogenic properties. The invention also relates to the DNA
XX CC sequences which encode the novel proteins. A wide variety of diseases are
XX CC the result of undesirable angiogenesis. The formation of new capillaries
XX CC from pre-existing vessels is essential for tumour growth and metastasis.
XX CC Basement membrane organisation is dependent on the assembly of a type IV
XX CC collagen network which may occur through the C-terminal globular non-
XX CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
XX CC forms are ubiquitously exhibited in human basement membranes. In the
XX CC present invention, cell surface receptors (in particular integrins) which
XX CC specifically bind anti-angiogenic proteins and peptides (in particular
XX CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
XX CC collagen) are disclosed. The proteins of the invention may inhibit tumour
XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in
XX CC endothelial cells and thus may exhibit cytostatic activity. The DNA
XX CC sequences of the invention may be useful in gene therapy. The present
XX CC sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
XX CC the "tumstatin" protein of the invention which was derived from the amino
XX CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This
XX CC sequence (Seq ID33) does not appear in the specification but was created
XX CC by the indexer from information given in the specification.
XX SQ Sequence 88 AA;
Query Match 32.8%; Score 80; DB 6; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.9e-74;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 45 FSLFLVQGNORAHGQDLGTLGSCLOQRTTTPFFLCNVNVCNPFASRNDYSYWLSTPALMP 104
DB 1 FSLFLVQGNORAHGQDLGTLGSCLOQRTTTPFFLCNVNVCNPFASRNDYSYWLSTPALMP 60
OY 105 MNMAPITGRALEPYISRCTV 124
DB 61 MNMAPITGRALEPYISRCTV 80

RESULT 31
AAU75600
ID AAU75600 standard; protein; 79 AA.
XX AC AAU75600;
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain mutant, Tum-5.
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX OS Homo sapiens.
XX PN WO200151523-A2.
XX PD 19-JUL-2001.
XX PF 08-JAN-2001; 2001WO-US000565.
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX PI WPI; 2002-188037/24.
XX DR
XX PT A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
XX PT treating disorders involving angiogenesis.
XX PS Example 40; Page: 205pp; English.
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI
XX CC domain, having one or more of the characteristics selected from: (a) the
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX CC proliferation of endothelial cells; and (c) the ability to cause
XX CC apoptosis of endothelial cells. Also described are the following: (1) use
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX CC analogue or allelic variant in the preparation of a medicament for
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX CC where the angiogenesis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,
XX CC alpha3, alpha5, alpha6, alphav, betav or beta3 subunit of integrin in the
XX CC preparation of a medicament for inhibiting angiogenesis or cell
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation
XX CC of a medicament for treating a proliferative disease in a vertebrate,
XX CC where the disease is characterised by angiogenesis that is mediated by
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and
XX CC (5) use of integrins in the preparation of a medicament for promoting or
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments
```

CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues
CC 54-132 of Tumstatin. Note: The present sequence is not shown in the
CC specification but is derived from the wild type human Tumstatin sequence
CC given in figure 18A (see AAU75599)
XX
SQ Sequence 79 AA;

Query Match 32.4%; Score 79; DB 5; Length 79;
Best Local Similarity 100.0%; Pred. No. 8.5e-73;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 NQAHGQDLGTLGSLQRFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITG 112
Db 1 NQAHGQDLGTLGSLQRFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITG 60

QY 113 RALEPYISRCTVCEGPAIA 131
Db 61 RALEPYISRCTVCEGPAIA 79

RESULT 32

ADA20264
ID ADA20264 standard; protein; 79 AA.

XX AC ADA20264;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-5 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cyostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX OS Homo sapiens.

XX PN WO200305257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 26; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV,
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cyostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of
CC the invention which was derived from the amino acid sequence of the alpha
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does
CC not appear in the specification but was created by the indexer from
CC information given in the specification.
XX
SQ Sequence 79 AA;

Query Match 32.4%; Score 79; DB 6; Length 79;
Best Local Similarity 100.0%; Pred. No. 8.5e-73;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDLGTLGSLQRFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGR 113
Db 1 QRAHGQDLGTLGSLQRFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGR 60

QY 114 ALEPYISRCTVCEGPAIAI 132

Db 61 ALEPYISRCTVCEGPAIAI 79

RESULT 33

AAU75599
ID AAU75599 standard; protein; 65 AA.

XX AC AAU75599;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tum-4.

XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and

XX treating disorders involving angiogenesis.

XX Example 36; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI

XX domain, having one or more of the characteristics selected from: (a) the

XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2,
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or
CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, Tum-4, which consists of residues
CC 181-244 of Tumstatin. Note: The present sequence is not shown in the
CC specification but is derived from the wild type human Tumstatin sequence
CC given in figure 18A (see AAU/5589)

XX SQ Sequence 65 AA;

Query Match 26.6%; Score 65; DB 5; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.6e-58;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 179 LEEFRAPFLECHGRGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQV 238
Db 1 LEEFRAPFLECHGRGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQV 60

Oy 239 CMKKR 243

Db 61 CMKKR 65

RESULT 34

ADA20263

ID ADA20263 standard; protein; 64 AA.

XX ADA20263;

AC ADA20263;

XX 20-NOV-2003 (first entry)

DT Human tumstatin deletion protein tum-4 amino acid sequence.

DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cyostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-4.

XX Homo sapiens.

OS Homo sapiens.

XX WO20003059257-A2.

FN

XX

PD

XX 24-JUL-2003.

PF 20-DEC-2002; 2002MO-US040938.

XX 21-DEC-2001; 2001US-00032221.

PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
growth, angiogenic activity or protein synthesis in a mammalian tissue.

PS Claim 94; SEQ ID NO 25; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
with anti-angiogenic properties. The invention also relates to the DNA
sequences which encode the novel proteins. A wide variety of diseases are
the result of undesirable angiogenesis. The formation of new capillaries
from pre-existing vessels is essential for tumour growth and metastasis.
Basement membrane organisation is dependent on the assembly of a type IV
collagen network which may occur through the C-terminal globular non-
collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
forms are ubiquitously exhibited in human basement membranes. In the
present invention, cell surface receptors (in particular integrins) which
specifically bind anti-angiogenic proteins and peptides (in particular
the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
collagen) are disclosed. The proteins of the invention may inhibit tumour
growth, angiogenic activity in mammalian tissue or protein synthesis in
endothelial cells and thus may exhibit cytostatic activity. The DNA
sequences of the invention may be useful in gene therapy. The present
sequence is that of tum-4, an abridged form of the "tumstatin" protein of
the invention which was derived from the amino acid sequence of the alpha
3 chain of human type IV collagen. Note: This sequence (Seq ID25) does
not appear in the specification but was created by the indexer from
CC information given in the specification.

XX SQ Sequence 64 AA;

Query Match 26.2%; Score 64; DB 6; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.7e-57;

Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQV 240

Db 1 EFRASPFLECHGRGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQV 60

Oy 241 KKRH 244

Db 61 KKRH 64

RESULT 35

AA95920

ID AA95920 standard; protein; 68 AA.

XX AC AA95920;

DT 20-NOV-2000 (first entry)

DE Human Goodpasture antigen Deltail/IV/V.

XX Goodpasture antigen; GPdeltail/IV/V; human; GPBP;

KW Goodpasture antigen binding protein; autoimmune disease; apoptosis;

XX cancer; tumour; therapy.

OS Homo sapiens.

XX WO200050607-A2.

FN

XX

XX (SAUS/) SAUS J.
 XX PI Saus J;
 XX WPI: 2000-572094/53.
 XX DR N-PSDB; AAA50368.
 XX
 PT Novel Goodpasture antigen binding proteins useful for diagnosing and
 PT treating autoimmune disorders, tumor, and preventing cell apoptosis.
 XX
 XX Claim 36; Page 153; 158pp; English.
 XX
 CC The present sequence is that of human recombinant Goodpasture antigen
 CC (GP) DeltaII, i.e. an alternative form of human GP resulting from
 CC splicing out of exon III. The recombinant protein, lacking the Met-1
 CC residue, was expressed in bacterial pellets using modified vector pET15b
 CC carrying GPdeltaII cDNA (see AAA50368). The provides novel Goodpasture
 CC antigen binding proteins (GPBPs, see AAY95900-11), which bind to and
 CC phosphorylate the unique N-terminal region of human GP, and which are
 CC highly expressed in several autoimmune conditions. Claimed methods for
 CC treating an autoimmune disorder, cell apoptosis or a tumour involve
 CC modifying the expression or activity of GPBP, especially using a GP-
 CC derived peptide, such as GPdeltaII
 XX
 XX Sequence 72 AA;
 Query Match 25.0%; Score 61; DB 3; Length 72;
 Best Local Similarity 100.0%; Pred. No. 2.3e-54;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 DB 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 QY 61 L 61
 DB 61 L 61
 RESULT 38
 AAY95921
 ID AAY95921 standard; protein; 72 AA.
 XX
 AC AAY95921;
 XX
 DT 20-NOV-2000 (first entry)
 XX
 DE Human Goodpasture antigen DeltaII/V.
 XX
 KW Goodpasture antigen; GPdeltaII/V; human; GPBP;
 KW goodpasture antigen binding protein; autoimmune disease; apoptosis;
 KW cancer; tumour; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO2000050607-A2.
 XX
 PD 31-AUG-2000.
 XX
 PF 24-FEB-2000; 2000WO-IB000324.
 XX
 PR 24-FEB-1999; 99US-0121483P.
 XX
 PA (SAUS/) SAUS J.
 XX
 PI Saus J;
 XX
 DR WPI: 2000-572094/53.
 DR N-PSDB; AAA50370.
 XX
 PT Novel Goodpasture antigen binding proteins useful for diagnosing and
 PT treating autoimmune disorders, tumor, and preventing cell apoptosis.

XX Claim 36; Page 155; 158pp; English.
 XX
 CC The present sequence is that of human recombinant Goodpasture antigen
 CC (GP) DeltaIII/V, i.e. an alternative form of human GP resulting from
 CC splicing out of exons III and V. The recombinant protein was expressed in
 CC bacterial pellets using modified vector pET15b carrying GPdeltaIII/V
 CC cDNA (see AAA50369). The provides Goodpasture antigen binding proteins
 CC (GPBPs, see AAY95900-11), which bind to and phosphorylate the unique N-
 CC terminal region of human GP, and which are highly expressed in several
 CC autoimmune conditions. Claimed methods for treating an autoimmune
 CC disorder, cell apoptosis or a tumour involve modifying the expression or
 CC activity of GPBP, especially using a GP-derived peptide, such as
 CC GPdeltaIII/V
 XX
 XX Sequence 72 AA;
 Query Match 25.0%; Score 61; DB 3; Length 72;
 Best Local Similarity 100.0%; Pred. No. 2.3e-54;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 DB 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 QY 61 L 61
 DB 61 L 61
 RESULT 39
 ABG79209
 ID ABG79209 standard; protein; 72 AA.
 XX
 AC ABG79209;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Human GP protein isoform GPdeltaIII.
 XX
 KW Goodpasture antigen binding protein; Goodpasture syndrome;
 KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
 KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
 KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
 KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
 KW pemphigoid; lichen planus; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200261430-A2.
 XX
 PD 08-AUG-2002.
 XX
 PF 31-JAN-2002; 2002WO-EP001010.
 XX
 PR 31-JAN-2001; 2001US-0265249P.
 XX
 PA (SAUS/) SAUS J.
 XX
 PI Saus J;
 XX
 DR WPI: 2002-619280/66.
 DR N-PSDB; ABS64492.
 XX
 PT Identifying candidate compounds for treating autoimmune conditions, e.g.
 PT Goodpasture syndrome or lupus, comprises identifying compounds that
 PT reduce phosphorylation of, or formation of conformational isomers of,
 PT target proteins.
 XX
 PS Example 3; Page 201; 217pp; English.
 XX
 CC The invention relates to identifying candidate compounds to treat an
 CC autoimmune condition by identifying compounds that reduce phosphorylation

CC of a first target protein (I) (which is selected from Goodpasture antigen
 CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
 CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
 CC Gln-Lys-Arg-Pro-Gln-Arg-His-Gly), or reduce formation of
 CC conformational isomers of the second target protein (II) (selected from
 CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
 CC NCI domain conformational isomer, which has an amino acid sequence
 CC identical to the wild type alpha3 type IV collagen NCI domain, is
 CC stabilised by disulphide bonds, and has a molecular weight in a non-
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
 CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)
 CC domain (also known as the GP antigen) or an MBP isoform
 XX
 SQ Sequence 72 AA;

Query Match 25.0%; Score 61; DB 5; Length 72;
 Best Local Similarity 100.0%; Pred. No. 2.3e-54;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GLKGRGDSGPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 Db 1 GLKGRGDSGPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 QY 61 L 61
 Db 61 L 61

RESULT 40

ABG79211
 ID ABG79211 standard; protein; 72 AA.

XX AC ABG79211;
 XX DT 15-NOV-2002 (first entry)
 XX DE Human GP protein isoform GPdeltaII/IV.

XX Goodpasture antigen binding protein; Goodpasture syndrome;
 KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
 KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
 KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
 KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
 KW pemphigoid; lichen planus; human.

XX OS Homo sapiens.

XX PN WO200261430-A2.

XX PD 08-AUG-2002.

XX PF 31-JAN-2002; 2002WO-EP001010.

XX PR 31-JAN-2001; 2001US-0265249P.

XX PA (SAUS/) SAUS J.

XX PI Saus J;

XX DR WPI; 2002-619280/66.

XX DR N-PSDB; ABS64494.

XX Identifying candidate compounds for treating autoimmune conditions, e.g.
 PT Goodpasture syndrome or lupus, comprises identifying compounds that
 PT reduce phosphorylation of, or formation of conformational isomers of,
 PT target proteins.

XX PS Disclosure; Page 203-204; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an
 CC autoimmune condition by identifying compounds that reduce phosphorylation
 CC of a first target protein (I) (which is selected from Goodpasture antigen
 CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
 CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
 CC Gln-Lys-Arg-Pro-Gln-Arg-His-Gly), or reduce formation of
 CC conformational isomers of the second target protein (II) (selected from
 CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
 CC NCI domain conformational isomer, which has an amino acid sequence
 CC identical to the wild type alpha3 type IV collagen NCI domain, is
 CC stabilised by disulphide bonds, and has a molecular weight in a non-
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
 CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)
 CC domain (also known as the GP antigen) or an MBP isoform
 XX

SQ Sequence 72 AA;

Query Match 25.0%; Score 61; DB 5; Length 72;
 Best Local Similarity 100.0%; Pred. No. 2.3e-54;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 Db 1 GLKGRGDSGPATWTRGTFVTRHSQTATPSCPEGTVPVLYSGFSFLVQGNQRAHQD 60
 QY 61 L 61
 Db 61 L 61

RESULT 41

ADD47061
 ID ADD47061 standard; protein; 230 AA.

XX AC ADD47061;

XX DT 29-JAN-2004 (first entry)

XX DE Rat Protein AAB72238, SEQ ID NO 12749.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX OS Rattus norvegicus.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEHO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI; 2003-268312/26.

XX DR GENBANK; AAB72238.

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 230 AA;

Query Match 16.0%; Score 39; DB 7; Length 230;
 Best Local Similarity 100.0%; Pred. NO. 2.5e-31;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYFSLASLNPERMFRKPIPTVKAG 227
 DB 175 ECHGRGTCNYNSYFSLASLNPERMFRKPIPTVKAG 213

RESULT 42
 AAY44171
 ID AAY44171 standard; protein; 471 AA.

XX
 AC AAY44171;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Bovine type IV collagen alpha3 chain protein.

XX Recombinant; bovine; alpha3 chain; type IV collagen; detection;
 XX Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.

XX Bos taurus.
 XX US5973120-A.
 XX 26-OCT-1999.
 XX
 PF 07-MAR-1995; 95US-00399889.
 XX
 PR 30-NOV-1990; 90US-00621091.
 XX
 PA (UYA) UNIV YALE.
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 PI Hudson BG, Reenders ST, Morrison KE;
 XX WPI; 1999-610317/52.

DR N-PSDB; AAZ28774.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for
 PT diagnosis and treatment of Goodpasture syndrome.

XX Claim 1; Col 31-34; 27pp; English.

XX This sequence represents a recombinant bovine alpha3 chain of type IV
 CC collagen polypeptide. The sequence corresponds to the 238 amino acids of
 CC the C-terminal end of the triple helical domain and all 233 amino acids
 CC of the C-terminal non-collagenous domain. Alpha3 chain collagen
 CC polypeptides are useful for detecting Goodpasture antibodies in blood or
 CC tissue from a human patient and for treating Goodpasture syndrome,
 CC especially by neutralising the antibodies in the blood. The polypeptides
 CC also have a nephrotrophic activity

XX Sequence 471 AA;

Query Match 16.0%; Score 39; DB 2; Length 471;
 Best Local Similarity 100.0%; Pred. NO. 4.7e-31;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALPEYISRCTVCEGPAIAIAVAHSQTDTIPPCP 145
 DB 334 MAPITGRALPEYISRCTVCEGPAIAIAVAHSQTDTIPPCP 372

RESULT 43

AAV56783
 ID AAY56783 standard; protein; 471 AA.

XX
 AC AAY56783;

XX 27-MAR-2000 (first entry)

XX Bovine alpha3 type IV collagen.

XX Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.

XX Bos sp.

XX US6007980-A.

XX 28-DEC-1999.

XX 07-OCT-1998; 98US-00167364.

XX 30-NOV-1990; 90US-00621091.

XX 07-MAR-1995; 95US-00399889.

XX (UNIV) UNIV KANSAS MEDICAL CENT.
 XX (UYA) UNIV YALE.

XX Hudson BG, Reenders ST, Morrison KE;

XX WPI; 2000-096371/08.

XX N-PSDB; AAZ46728.

XX Diagnosing and treating Goodpasture syndrome using a peptide derived from
 PT type IV collagen.

XX Disclosure; Col 19-24; 26pp; English.

XX The invention provides a method of detecting Goodpasture antibodies in
 CC the fluid of a patient by contacting it with a peptide comprising at most
 CC 218 amino acids of the human alpha3 chain type IV collagen that contains
 CC the fragment shown in AAY56785. The methods are useful for the diagnosis
 CC and treatment of Goodpasture syndrome. The present sequence represents
 CC the bovine alpha3 chain of type IV collagen

XX Sequence 471 AA;

Query Match 16.0%; Score 39; DB 3; Length 471;

Best Local Similarity 100.0%; Pred. No. 4.7e-31; Mismatches 0; Indels 0; Gaps 0;
Matches 39; Conservative 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 372

RESULT 44
AAE09483
ID AAE09483 standard; protein; 471 AA.
XX
AC AAE09483;
XX
DT 19-NOV-2001 (first entry)
XX
DE Bovine alpha-3 chain of type IV collagen protein.
XX
KW Bovine; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome.
XX
OS Bos taurus.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
XX
XX 07-MAR-1995; 95US-00399889.
XX
XX 07-OCT-1998; 98US-00167364.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX WPI; 2001-540401/60.
XX
XX N-PSDB; AAD16399.
XX
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
XX Goodpasture antibodies from bodily fluid/tissue from patient or for
XX treating Goodpasture syndrome by contacting bodily fluid or tissue with
XX the polypeptide.
XX
XX Example 4; Col 33-36; 46pp; English.
XX
XX The invention relates to a method for detecting Goodpasture antibodies
XX from a bodily fluid or tissue of a patient. The method comprises
XX contacting the bodily fluid or tissue with alpha-3 chain type (IV)
XX collagen polypeptide that contains a conformational epitope for the
XX Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
XX detecting Goodpasture antibodies from a bodily fluid or tissue from a
XX patient, and for treating Goodpasture syndrome in a patient. The present
XX sequence is bovine alpha-3 chain of type IV collagen protein
XX
XX Sequence 471 AA;
XX
Query Match 16.0%; Score 39; DB 4; Length 471;
Best Local Similarity 100.0%; Pred. No. 4.7e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 372

RESULT 45
ABG79213
ID ABG79213 standard; peptide; 72 AA.
XX
XX ABG79213;
XX

DT 15-NOV-2002 (first entry)
XX
DE Human GP protein isoform GPdeltaIII phosphorylation site region.
XX
KW Goodpasture antigen binding protein; Goodpasture syndrome;
KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
KW pemphigoid; lichen planus; human.
XX
XX Homo sapiens.
XX
XX WO200261430-A2.
XX
XX 08-AUG-2002.
XX
XX 31-JAN-2002; 2002WO-EP001010.
XX
XX 31-JAN-2001; 2001US-0265249P.
XX
XX (SAUS/) SAUS J.
XX
XX Saus J;
XX
XX WPI; 2002-619280/66.
XX
XX Identifying candidate compounds for treating autoimmune conditions, e.g.
XX Goodpasture syndrome or lupus, comprises identifying compounds that
XX reduce phosphorylation of, or formation of conformational isomers of,
XX target proteins.
XX
XX Example 3; Fig 17; 217pp; English.
XX
XX The invention relates to identifying candidate compounds to treat an
XX autoimmune condition by identifying compounds that reduce phosphorylation
XX of a first target protein (I) (which is selected from Goodpasture antigen
XX binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)
XX domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
XX Ala-Thr-Tyr-Thr-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
XX Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
XX conformational isomers of the second target protein (II) (selected from
XX an alpha3 type IV collagen NCI domain polypeptide and myelin basic
XX protein, MBP). Also included are (1) an isolated type IV collagen alpha3
XX NCI domain conformational isomer, which has an amino acid sequence
XX identical to the wild type alpha3 type IV collagen NCI domain, is
XX stabilised by disulphide bonds, and has a molecular weight in a non-
XX reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
XX type IV collagen alpha3 NCI domain. The human gene for GPBP is located on
XX chromosome 5q13. The method is useful for treating autoimmune conditions,
XX such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
XX lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
XX sequence represents an alpha3 type IV collagen non-collagenous (NCI)
XX domain (also known as the GP antigen) or an MBP phosphorylation site
XX
XX Sequence 72 AA;
XX
Query Match 15.2%; Score 37; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 9.7e-30;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKRGDSGPATWTRGTFVTRHSQTTAIPSCPEG 37
DB 1 GLKGKRGDSGPATWTRGTFVTRHSQTTAIPSCPEG 37

RESULT 46
AAE09503
ID AAE09503 standard; peptide; 36 AA.
XX
XX AAE09503;
XX

```
DT 19-NOV-2001 (first entry)
XX Human C8 alpha-3 peptide to construct alpha1/alpha3(IV)NC1 protein.
DE
XX
DE Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome; C8 alpha-3 peptide.
XX
XX Homo sapiens.
OS
XX US6277558-B1.
PN
XX 21-AUG-2001.
PD
XX 12-NOV-1999; 99US-00439897.
PF
XX 30-NOV-1990; 90US-00621091.
PR 07-MAR-1995; 95US-00399889.
PR 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA
XX Hudson BG;
PI
XX WPI; 2001-540401/60.
DR
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
PT Goodpasture antibodies from bodily fluid/tissue from patient or for
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with
PT the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
PS
XX The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is human alpha chain peptide used for constructing human
CC alpha1/alpha3 (IV) NC1 fusion protein
XX
XX Sequence 36 AA;
SQ
Query Match 14.8%; Score 36; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 5.5e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 209 SLNPFMRKPISTVKAGELEKIIIRQCQVCMKKRH 244
DB 1 SLNPFMRKPISTVKAGELEKIIIRQCQVCMKKRH 36
RESULT 47
AAR79163
ID AAR79163 standard; protein; 471 AA.
XX
XX AAR79163;
AC
XX 22-DEC-1995 (first entry)
DT
XX Partial sequence of bovine alpha 3 chain of type IV collagen.
DE
XX Type IV collagen; alpha 3 chain.
KW Bos taurus.
XX
XX US5424408-A.
PN
XX 13-JUN-1995.
PD
XX 30-NOV-1990; 90US-00621091.
XX
XX
```

```
PR 30-NOV-1990; 90US-00621091.
XX
XX (UYIA ) UNIV YALE.
PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Morrison KE, Reenders ST, Hudson BG;
XX
XX WPI; 1995-262631/34.
DR N-PSDB; AAQ96290.
DR
XX CDNA's encoding human or bovine alpha-3 type 4 collagen peptide(s) -
PT useful for detection and therapeutic removal of antibodies associated
PT with Goodpasture syndrome.
XX
XX Disclosure; Col 5-8; 33pp; English.
PS
XX Using the PCR with primers derived from each end of the known 27 AA
CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp bovine
CC genomic fragment was amplified. This fragment was then used to a bovine
CC lens cDNA library and a 1.5 kb partial cDNA clone was obt'd. (clone
CC KMC15). This encodes 238 residues of the triple helical collagenous
CC domain and all 233 residues of the C-terminal non-collagenous (NC1)
CC domain of the alpha 3 (IV) chain. An isolated and substantially pure nt.
CC having the sequence in AAQ96290 is claimed
XX
XX Sequence 471 AA;
SQ
Query Match 12.3%; Score 30; DB 2; Length 471;
Best Local Similarity 100.0%; Pred. No. 8.1e-22;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 116 EPIYSRCTVCEGPAIAIAVHSQTTDIPPCP 145
DB 343 EPIYSRCTVCEGPAIAIAVHSQTTDIPPCP 372
RESULT 48
AAB09501
ID AAB09501 standard; peptide; 26 AA.
XX
XX AAB09501;
AC
XX 19-NOV-2001 (first entry)
DT
XX Human C7 alpha-3 peptide to construct alpha1/alpha3(IV)NC1 protein.
DE
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome; C7 alpha-3 peptide.
XX
XX Homo sapiens.
OS
XX US6277558-B1.
PN
XX 21-AUG-2001.
PD
XX 12-NOV-1999; 99US-00439897.
PF
XX 30-NOV-1990; 90US-00621091.
PR 07-MAR-1995; 95US-00399889.
PR 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA
XX Hudson BG;
PI
XX WPI; 2001-540401/60.
DR
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
PT Goodpasture antibodies from bodily fluid/tissue from patient or for
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with
PT the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
PS
```


CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is the amino acid sequence of the 17 peptide of the invention,
CC derived from the amino acid sequence of tumstatin, which in turn was
CC derived from the amino acid sequence of human type IV collagen alpha 3
CC chain.

XX SQ Sequence 25 AA;

Query Match 10.2%; Score 25; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 7.7e-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMTPLFCNVNDVCFASRNDYSYL 97
Db 1 TMTPLFCNVNDVCFASRNDYSYL 25
|||||

RESULT 51

ADCI17661
ID ADCI17661 standard; peptide; 22 AA.

XX AC ADCI17661;

XX DT 18-DEC-2003 (first entry)

XX Type IV collagen NC1 domain related peptide SEQ ID NO:266.

DE crystallised NC1 domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
KW tumour metastasis inhibitor; tumour growth inhibitor;
KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.

XX Synthetic.

OS Homo sapiens.

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

XX 29-OCT-2001; 2001US-0351289P.

XX 22-MAR-2002; 2002US-0366854P.

XX 03-JUN-2002; 2002US-0385362P.

XX (UNIV.) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT basal lamina membrane formation in cell or tissue development.

XX Claim 57; SEQ ID NO 266; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a

CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (5) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NC1
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen (I) has cytostatic,
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and
CC antiulcer activities, and can be used as an inhibitor of angiogenesis;
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
CC invention.

XX SQ Sequence 22 AA;

Query Match 9.0%; Score 22; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 8.2e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTTMPFLFCNVNDVCFASRND 92

Db 1 FTTMPFLFCNVNDVCFASRND 22
|||||

RESULT 52

ADCI17414

ID ADCI17414 standard; peptide; 22 AA.

XX AC ADCI17414;

XX DT 18-DEC-2003 (first entry)

XX Type IV collagen NC1 domain related peptide SEQ ID NO:15.

DE crystallised NC1 domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
KW tumour metastasis inhibitor; tumour growth inhibitor;
KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 1..5 /note= "optionally serially deleted"

XX Misc-difference 18..22 /note= "optionally serially deleted"

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

XX 29-OCT-2001; 2001US-0351289P.


```

PR 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAWOORTHY M.
PA (HUDS/) HUDSON B.
XX Sundaramoorthy M, Hudson B;
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Claim 11; SEQ ID NO 15; 168pp; English.
XX
XX The present invention describes a crystallised NCL domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCL
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCL domain hexamer of type IV collagen (I) has cytostatic,
XX antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and
XX antitumor activities, and can be used as an inhibitor of angiogenesis,
XX tumor growth, tumor metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX SQ Sequence 22 AA;
XX
XX Query Match 9.0%; Score 22; DB 7; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 8.2e-15;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 161 EFRASPFLCHGRCNYNS 202
XX Db 1 EFRASPFLCHGRCNYNS 22
XX
XX RESULT 53
XX AAY95912
XX ID AAY95912 standard; peptide; 21 AA.
XX AC AAY95912;
XX XX
XX 20-NOV-2000 (first entry)
XX DT
XX DE Human Goodpasture antigen N-terminal peptide GPpepl.
XX
XX Goodpasture antigen binding protein; Goodpasture syndrome; antigen;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX OS
XX Homo sapiens.
XX OS
XX WO200261430-A2.
XX FN
XX 08-AUG-2002.
XX PD
XX 31-JAN-2002; 2002WO-EP001010.
XX PF
XX 31-JAN-2001; 2001US-0265249P.
XX PR
XX (SAUS/) SAUS J.
XX PA
XX Saus J;
XX PI
XX WPI; 2002-619280/66.
XX DR
XX
XX

```

```

PF 24-FEB-2000; 2000WO-IB000324.
XX
XX 24-FEB-1999; 99US-0121483P.
XX
XX (SAUS/) SAUS J.
XX
XX Saus J;
XX
XX WPI; 2000-572094/53.
XX
XX Novel Goodpasture antigen binding proteins useful for diagnosing and
XX treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX
XX Example 1; Page 21; 158pp; English.
XX
XX The present sequence is that of GPpepl, the N-terminal 21 amino acids of
XX human Goodpasture antigen (GP). The peptide was used to search for
XX proteins that interacted with it, leading to the identification of human
XX Goodpasture binding protein (GPBP, see AAY95900), a novel
XX serine/threonine kinase that specifically binds to and phosphorylates
XX GPpepl. GPpepl was also used to characterise the phosphorylation activity
XX of GPBP. The invention provides nucleic acids (see AAA50341-53) encoding
XX GPBP, recombinant vectors, host cells, encoded polypeptides (see AAY95900
XX -11) and antibodies. It also provides methods for detecting the presence
XX of an autoimmune condition or apoptosis by detecting an increase in GPBP
XX expression, and methods for treating an autoimmune disorder, apoptosis or
XX a tumour by modifying GPBP expression or activity, especially using a GP-
XX derived peptide
XX
XX SQ Sequence 21 AA;
XX
XX Query Match 8.6%; Score 21; DB 3; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-14;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 KGRGDSGSPATWTRGFVFT 23
XX Db 1 KGRGDSGSPATWTRGFVFT 21
XX
XX RESULT 54
XX ABG79202
XX ID ABG79202 standard; peptide; 21 AA.
XX
XX AC ABG79202;
XX
XX 15-NOV-2002 (first entry)
XX DT
XX DE Human Goodpasture protein peptide, GPpepl.
XX
XX Goodpasture antigen binding protein; Goodpasture syndrome; antigen;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX OS
XX Homo sapiens.
XX OS
XX WO200261430-A2.
XX FN
XX 08-AUG-2002.
XX PD
XX 31-JAN-2002; 2002WO-EP001010.
XX PF
XX 31-JAN-2001; 2001US-0265249P.
XX PR
XX (SAUS/) SAUS J.
XX PA
XX Saus J;
XX PI
XX WPI; 2002-619280/66.
XX DR
XX
XX

```

PT Identifying candidate compounds for treating autoimmune conditions, e.g.
 PT Goodpasture syndrome or lupus, comprises identifying compounds that
 PT reduce phosphorylation of, or formation of conformational isomers of,
 PT target proteins.

XX Claim 1; Page 26; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an
 CC autoimmune condition by identifying compounds that reduce phosphorylation
 CC of a first target protein (I) (which is selected from Goodpasture antigen
 CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCL)
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-
 CC Ala-Thr-Trip-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
 CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
 CC conformational isomers of the second target protein (II) (selected from
 CC an alpha3 type IV collagen NCL domain polypeptide and myelin basic
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
 CC NCL domain conformational isomer, which has an amino acid sequence
 CC identical to the wild type alpha3 type IV collagen NCL domain, is
 CC stabilised by disulphide bonds, and has a molecular weight in a non-
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
 CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
 CC type IV collagen alpha3 NCL domain. The human gene for GPBP is located on
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCL)
 CC domain (also known as the GP antigen) peptide antigen

XX Sequence 21 AA;

Query Match 8.6%; Score 21; DB 5; Length 21;
 Best Local Similarity 100.0%; Pred. No. 8.3e-14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KGRGDSGSPATWTRGFVET 23
 DB 1 KGRGDSGSPATWTRGFVET 21
 |||||

RESULT 55

ADCL17642
 ID ADC17642 standard; peptide; 21 AA.

XX
 AC ADC17642;

XX
 DT 18-DEC-2003 (first entry)

XX Type IV collagen NCL domain related peptide SEQ ID NO:247.

XX crystallised NCL domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cyrostatic; antiapoptotic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

XX Synthetic.

OS Homo sapiens.

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

PR 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.

PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

PI Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.

XX Disclosure; SEQ ID NO 247; 168pp; English.

XX The present invention describes a crystallised NCL domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (5) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NCL
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NCL domain hexamer of type IV collagen (I) has cytostatic,
 CC antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic, and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.

XX Sequence 21 AA;

Query Match 8.6%; Score 21; DB 7; Length 21;
 Best Local Similarity 100.0%; Pred. No. 8.3e-14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 FWLASLNPERMFRKPIPISTVK 225

DB 1 FWLASLNPERMFRKPIPISTVK 21
 |||||

RESULT 56

AAU75604

ID AAU75604 standard; peptide; 20 AA.

XX AAU75604;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain, Tumstatin, mutant T4.

XX Human; type IV collagen alpha 3 chain; cyrostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NCL domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

```
XX PF 08-JAN-2001; 2001WO-US000565.
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PA Kalluri R;
XX PI WPI; 2002-188037/24.
XX DR A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX PT treating disorders involving angiogenesis.
XX PS Example 40; Page 133; 205pp; English.
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX CC domain, having one or more of the characteristics selected from: (a) the
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX CC proliferation of endothelial cells; and (c) the ability to cause
XX CC apoptosis of endothelial cells. Also described are the following: (1) use
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX CC analogue or allelic variant in the preparation of a medicament for
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX CC where the angiogenesis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,
XX CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
XX CC preparation of a medicament for inhibiting angiogenesis or cell
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation
XX CC of a medicament for treating a proliferative disease in a vertebrate,
XX CC where the disease is characterised by angiogenesis that is mediated by
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and
XX CC (5) use of integrins in the preparation of a medicament for promoting or
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments
XX CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX CC or allelic variants are useful in the preparation of a medicament for
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where
XX CC the angiogenesis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits; or by promoting or
XX CC inducing endothelial cell apoptosis in a tissue, where the endothelial
XX CC cell apoptosis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits. The medicament is useful
XX CC in inhibiting tumour growth and for the regression of an established
XX CC tumour. The present sequence represents the amino acid sequence of human
XX CC type IV collagen alpha 3 chain mutant, T4, which consists of residues 84-
XX CC 103 of Tumstatin
XX SQ Sequence 20 AA;
XX
XX Query Match 8.2%; Score 20; DB 5; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 8.5e-13; Indels 0; Gaps 0;
XX Matches 20; Conservative 0; Mismatches 0;
XX
XX QY 83 DVCNPARNDYSYWLSTPAL 102
XX |||||
XX Db 1 DVCNPARNDYSYWLSTPAL 20
XX
XX RESULT 57
XX AAU75602
XX ID AAU75602 standard; peptide; 20 AA.
XX XX
XX AC AAU75602;
```

```
XX DT 08-MAY-2002 (first entry)
XX DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T2.
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX KW Tumstatin; angiogenesis; tumour; mutant; mutant.
XX OS Homo sapiens.
XX PN WO200151523-A2.
XX PD 19-JUL-2001.
XX PF 08-JAN-2001; 2001WO-US000565.
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2002-188037/24.
XX CC A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX CC treating disorders involving angiogenesis.
XX CC Example 40; Page 133; 205pp; English.
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX CC domain, having one or more of the characteristics selected from: (a) the
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX CC proliferation of endothelial cells; and (c) the ability to cause
XX CC apoptosis of endothelial cells. Also described are the following: (1) use
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX CC analogue or allelic variant in the preparation of a medicament for
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX CC where the angiogenesis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,
XX CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
XX CC preparation of a medicament for inhibiting angiogenesis or cell
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation
XX CC of a medicament for treating a proliferative disease in a vertebrate,
XX CC where the disease is characterised by angiogenesis that is mediated by
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and
XX CC (5) use of integrins in the preparation of a medicament for promoting or
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments
XX CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX CC or allelic variants are useful in the preparation of a medicament for
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where
XX CC the angiogenesis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits; or by promoting or
XX CC inducing endothelial cell apoptosis in a tissue, where the endothelial
XX CC cell apoptosis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits. The medicament is useful
XX CC in inhibiting tumour growth and for the regression of an established
XX CC tumour. The present sequence represents the amino acid sequence of human
XX CC type IV collagen alpha 3 chain mutant, T2, which consists of residues 54-
XX CC 73 of Tumstatin
XX SQ Sequence 20 AA;
```

Query Match 8.2%; Score 20; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 NORAHGQDLTGSLQRT 72
DB 1 NORAHGQDLTGSLQRT 20
|||||

RESULT 58
ID AAU75603 standard; peptide; 20 AA.
XX AAU75603;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T3.
XX
KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutain; mutant.
XX
OS Homo sapiens.
XX
PN WO200151523-A2.
XX
PD 19-JUL-2001.
XX
PF 08-JAN-2001; 2001WO-US000565.
XX
PR 07-JAN-2000; 2000US-00479118.
PR 04-APR-2000; 2000US-00543371.
PR 21-JUL-2000; 2000US-00625191.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
PF WPI; 2002-188037/24.
XX
PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
PT treating disorders involving angiogenesis.
XX
PS Claim 33; Page 153; 205pp; English.

CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
CC domain, having one or more of the characteristics selected from: (a) the
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
CC proliferation of endothelial cells; and (c) the ability to cause
CC apoptosis of endothelial cells. Also described are the following: (1) use
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
CC analogue or allelic variant in the preparation of a medicament for
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
CC where the angiogenesis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; or (b) by
CC promoting or inducing endothelial cell apoptosis in a tissue, where the
CC endothelial cell apoptosis is mediated by one or more endothelial cell
CC integrins or one or more endothelial cell integrin subunits; (2) use of
CC an antibody or peptide that specifically binds the alpha1, alpha2, or
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin in the
CC preparation of a medicament for inhibiting angiogenesis or cell
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
CC fragment or peptide of receptor-mediated angiogenesis in the preparation
CC of a medicament for treating a proliferative disease in a vertebrate,
CC where the disease is characterised by angiogenesis that is mediated by
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
CC the presence of a medicament for promoting angiogenesis in a tissue; and
CC (5) use of integrins in the preparation of a medicament for promoting or

CC inducing angiogenesis or cell proliferation in a tissue. The fragments
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
CC or allelic variants are useful in the preparation of a medicament for
CC treating a disorder involving inhibiting angiogenesis in a tissue, where
CC the angiogenesis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits; or by promoting or
CC inducing endothelial cell apoptosis in a tissue, where the endothelial
CC cell apoptosis is mediated by one or more endothelial cell integrins or
CC one or more endothelial cell integrin subunits. The medicament is useful
CC in inhibiting tumour growth and for the regression of an established
CC tumour. The present sequence represents the amino acid sequence of human
CC type IV collagen alpha 3 chain mutant, T3, which consists of residues 69-
CC 88 of Tumstatin
XX
SQ Sequence 20 AA;
Query Match 8.2%; Score 20; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LQRTTTPFLFCNVNVCNF 87
DB 1 LQRTTTPFLFCNVNVCNF 20
|||||

RESULT 59
ADA20267
ID ADA20267 standard; peptide; 20 AA.
XX
AC ADA20267;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human tumstatin deletion peptide T3 amino acid sequence.
XX
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T3.
XX
OS Homo sapiens.
XX
PN WO2003059257-A2.
XX
PD 24-JUL-2003.
XX
PF 20-DEC-2002; 2002WO-US040938.
XX
PR 21-DEC-2001; 2001US-00032221.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
DR WPI; 2003-587256/55.
DR N-PSDB; ADA20224.
XX
PT New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Claim 94; Page 131; 240pp; English.

CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of peptide T3, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen.
XX
XX Sequence 20 AA;

Query Match 8.2%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 LORFTTTPFLFCNVNDCNF 87
Db 1 LORFTTTPFLFCNVNDCNF 20

RESULT 60
ADA20266
ID ADA20266 standard; peptide; 20 AA.

XX ADA20266;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion peptide T2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T2.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Example 40; Page 131; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of peptide T2, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen.
XX
XX Sequence 20 AA;

Query Match 8.2%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 NORAHGQDLTGSCICLQRF 72
Db 1 NORAHGQDLTGSCICLQRF 20

RESULT 61
ADA20268
ID ADA20268 standard; peptide; 20 AA.

XX ADA20268;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion peptide T4 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T4.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; Page 131; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of peptide T4, an abridged form of the "tumstatin"
XX protein of the invention which was derived from the amino acid sequence

CC of the alpha 3 chain of human type IV collagen.
 XX Sequence 20 AA;
 SQ

Query Match 8.2%; Score 20; DB 6; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.5e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 DVCNFSARNDSYWLSTPAL 102
 Db 1 DVCNFSARNDSYWLSTPAL 20
 |||||
 |||||

RESULT 62
 ADCL17684
 ID ADCL17684 standard; peptide; 20 AA.
 XX
 AC ADCL17684;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Type IV collagen NC1 domain related peptide SEQ ID NO:289.
 XX
 KW crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antiapoptotic; antiproliferative;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO2003012122-A2.
 XX
 PD 13-FEB-2003.
 XX
 PF 26-JUL-2002; 2002WO-US023763.
 XX
 PR 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX
 PI Sundaramoorthy M, Hudson B;
 XX
 DR WPI; 2003-332730/31.
 XX
 PT New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 PS Claim 57; SEQ ID NO 289; 168pp; English.
 XX

CC The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,

CC antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 20 AA;

Query Match 8.2%; Score 20; DB 7; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.5e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LQRFMTMPFLFCNVNDVCF 87
 Db 1 LQRFMTMPFLFCNVNDVCF 20
 |||||
 |||||

RESULT 63
 AAU75606
 ID AAU75606 standard; peptide; 19 AA.
 XX
 AC AAU75606;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T6.
 XX
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutant.
 XX
 OS Homo sapiens.
 XX
 PN WO200151523-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 08-JAN-2001; 2001WO-US000565.
 XX
 PR 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX
 PI Kalluri R;
 XX
 DR WPI; 2002-188037/24.
 XX
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.
 XX
 PS Example 40; Page 133; 205pp; English.
 XX

CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, T6, which consists of residues 114
 CC -132 of Tumstatin
 XX
 SQ Sequence 19 AA;

Query Match 7.8%; Score 19; DB 5; Length 19;
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPYISRCTVCEGPAIA 131
 DB 1 RALEPYISRCTVCEGPAIA 19
 |||||

RESULT 64
 AAU75605
 ID AAU75605 standard; peptide; 19 AA.
 AC AAU75605;

DT 08-MAY-2002 (first entry)
 DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T5.
 XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.
 XX WO200151523-A2.
 PN 19-JUL-2001.
 XX 08-JAN-2001; 2001WO-US000565.
 PF 07-JAN-2000; 2000US-00479118.
 PR 04-APR-2000; 2000US-00543371.
 PR 21-JUL-2000; 2000US-00625191.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA Kalluri R;
 XX WPI; 2002-188037/24.

XX
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.
 XX
 PS Example 40; Page 133; 205pp; English.

CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 3 chain mutant, T5, which consists of residues 99-
 CC 117 of Tumstatin
 XX
 SQ Sequence 19 AA;

Query Match 7.8%; Score 19; DB 5; Length 19;
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 STPALPMNNAPITGRALE 116
 DB 1 STPALPMNNAPITGRALE 19
 |||||

RESULT 65
 ABP58053
 ID ABP58053 standard; peptide; 19 AA.
 XX
 AC ABP58053;

DT 03-MAR-2003 (first entry)
 XX Collagen type IV alpha3 chain noncollagenous 1 domain peptide.
 DE Angiogenesis; inhibitor; collagen; cytostatic; antiinflammatory;
 XX immunosuppressive; antiarthritic; antiarteriosclerotic; osteopathic;
 KW antirheumatic; ophthalmological.
 XX Homo sapiens.
 OS
 XX WO200266512-A1.

XX 29-AUG-2002.
 PD 15-FEB-2002; 2002WO-US005211.
 XX 16-FEB-2001; 2001US-0269537P.
 XX 14-SEP-2001; 2001US-0322047P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA Scialdone MA, Mousa SA, Shuey SW;
 PI WPI; 2003-111767/10.
 XX
 XX New angiogenesis-inhibitory tripeptide useful for inhibiting endothelial
 PT cell tube formation in angiogenesis-dependent diseases such as cancer,
 PT ocular neovascularization and inflammatory diseases.
 XX
 PS Disclosure; Page 3; 48pp; English.
 XX
 CC The present sequence is that of amino acid residues 185-203 of the
 CC noncollagenous 1 (NC1) domain of the alpha3-chain of basement membrane
 CC collagen type IV. A peptide comprising these residues promoted adhesion
 CC of human melanoma cells by 50-60% over controls and inhibited their
 CC proliferation by 40%. Alanine substitutions through the peptide sequence
 CC indicated that the observed activities were dependent on the presence of
 CC residues 189-191, referred to as the SNS (Ser-Asn-Ser) sequence. The
 CC invention provides methods and compositions for inhibiting endothelial
 CC cell tube formation, the initial step of tumour angiogenesis.
 CC Tripeptides, preferably SNS or SQS (Ser-Gln-Ser) tripeptides, are used to
 CC inhibit angiogenesis-mediated processes such as ocular neovascular
 CC diseases, choroidal neovascular diseases, retina neovascular diseases,
 CC neovascularization of the angle, Bartonellosis, chronic inflammation,
 CC osteoarthritis, rheumatoid arthritis, atherosclerosis phemphigoid,
 CC trachoma, or Osler-Webber-Rendu disease (all claimed). They are also
 CC useful for treating cancer, inflammatory disorders and autoimmune
 CC diseases
 XX
 SQ Sequence 19 AA;
 XX
 Query Match 7.8%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 196 CNYNSNSYFWLASLNPER 214
 DB 1 CNYNSNSYFWLASLNPER 19
 XX
 RESULT 66
 ADA20269
 ID ADA20269 standard; peptide; 19 AA.
 XX
 AC ADA20269;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumstatin deletion peptide T5 amino acid sequence.
 XX
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytotatic; gene therapy; alpha 3 chain; tumstatin; human; T5.
 XX
 OS Homo sapiens.
 XX
 PN WO2003059257-A2.
 XX
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PF (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PR 21-DEC-2001; 2001US-00032221.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX Kalluri R;
 XX WPI; 2003-587256/55.
 DR N-PSDB; ADA20224.
 XX
 XX New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 PT
 XX Example 40; Page 131; 240pp; English.
 XX
 CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins), which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is that of peptide T5, an abridged form of the "tumstatin"
 CC protein of the invention which was derived from the amino acid sequence
 CC of the alpha 3 chain of human type IV collagen.
 XX
 SQ Sequence 19 AA;
 XX
 Query Match 7.8%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 98 STPALMPMNMAPITGRALE 116
 DB 1 STPALMPMNMAPITGRALE 19
 XX
 RESULT 67
 ADA20265
 ID ADA20265 standard; peptide; 19 AA.
 XX
 AC ADA20265;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumstatin deletion peptide T1 amino acid sequence.
 XX
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytotatic; gene therapy; alpha 3 chain; tumstatin; human; T1.
 XX
 OS Homo sapiens.
 XX
 PN WO2003059257-A2.
 XX
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PF (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PI Kalluri R;
XX WPI; 2003-587256/55.
DR N-PSDB; ADA20224.
XX
PT New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Example 40; Page 131; 240pp; English.
XX
CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of peptide T1, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen.
XX
SQ Sequence 19 AA;

Query Match 7.8%; Score 19; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.6e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTTRG 19
DB 1 GLKGRGDSGSPATWTTRG 19

RESULT 68
ADA20270
ID ADA20270 standard; peptide; 19 AA.

AC ADA20270;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human tumstatin deletion peptide T6 amino acid sequence.
XX
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T6.
XX
OS Homo sapiens.
XX
PN WO2003059257-A2.
XX
PD 24-JUL-2003.
XX
PF 20-DEC-2002; 2002WO-US040938.
XX
PR 21-DEC-2001; 2001US-00032221.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
XX WPI; 2003-587256/55.
DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
PS Example 40; Page 131; 240pp; English.
XX
CC This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is that of peptide T6, an abridged form of the "tumstatin"
CC protein of the invention which was derived from the amino acid sequence
CC of the alpha 3 chain of human type IV collagen.
XX
SQ Sequence 19 AA;

Query Match 7.8%; Score 19; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.6e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPYISRCTVCEGPAIA 131
DB 1 RALEPYISRCTVCEGPAIA 19

RESULT 69
AAU75601
ID AAU75601 standard; peptide; 20 AA.

AC AAU75601;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T1.
XX
KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
OS Homo sapiens.
XX
PN WO200151523-A2.
XX
PD 19-JUL-2001.
XX
PF 08-JAN-2001; 2001WO-US000565.
XX
PR 07-JAN-2000; 2000US-00479118.
PR 04-APR-2000; 2000US-00543371.
PR 21-JUL-2000; 2000US-00623191.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
XX WPI; 2002-188037/24.
XX
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
PT treating disorders involving angiogenesis.
XX

Example 40; Page 133; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, or alpha3, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors are inhibited by Arresten, Canstatin or Tumstatin; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 3 chain mutant, T1, which consists of residues 1-20 of Tumstatin

Sequence 20 AA;

Query Match 7.8%; Score 19; DB 5; Length 20;

Best Local Similarity 100.0%; Pred. No. 9e-12; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLXGKRGDSGSPATWTRG 19

Db 2 GLXGKRGDSGSPATWTRG 20

RESULT 70

ADCL17655

ID ADCL17655 standard; peptide; 18 AA.

AC ADCL17655;

XX 18-DEC-2003 (first entry)

XX Type IV collagen NC1 domain related peptide SEQ ID NO:260.

XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
 KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

OS Synthetic.
 OS Homo sapiens.

PN WO2003012122-A2.

PD 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

PR 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.

XX 03-JUN-2002; 2002US-0385362P.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or

PT condition consisting of glaucoma or blood-borne tumors or for inhibiting

PT basal lamina membrane formation in cell or tissue development.

XX Claim 57; SEQ ID NO 260; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a pharmaceutical composition comprising the polypeptide and a carrier; (3) inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated disease or condition in a mammal; (5) inhibiting tumour metastasis or growth; (6) inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue; (7) inhibiting basal lamina membrane formation in cell or tissue development; (8) identifying inhibitors of type IV collagen assembly; and (9) an inhibitor of type IV collagen assembly. A crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic, antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and antiulcer activities, and can be used as an inhibitor of angiogenesis, tumour growth, tumour metastasis, endothelial cell adhesion, endothelial cell proliferation, and basal lamina assembly. A (I) polypeptide can be used for treating an angiogenesis-mediated disease or condition consisting of glaucoma, sickle cell anaemia, ulcerative colitis, psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours or for inhibiting basal lamina membrane formation in cell or tissue development. The methods are useful for inhibiting angiogenesis in tissue, inhibiting tumour metastasis or growth, inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue, and identifying inhibitors of type IV collagen assembly. The present sequence represents a peptide which is used in the exemplification of the present invention.

XX Sequence 18 AA;

Query Match 7.4%; Score 18; DB 7; Length 18;

Best Local Similarity 100.0%; Pred. No. 8.7e-11;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PFLFCNVNDVCFASRND 92

Db 1 PFLFCNVNDVCFASRND 18

RESULT 71

ADCL17672

ID ADCL17672 standard; peptide; 18 AA.

XX ADCL17672;

XX 18-DEC-2003 (first entry)

XX

DE Type IV collagen NC1 domain related peptide SEQ ID NO:277.
 XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO2003012122-A2.
 XX
 PD 13-FEB-2003.
 XX
 PD 26-JUL-2002; 2002WO-US023763.
 XX
 PR 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX
 PI Sundaramoorthy M, Hudson B;
 XX
 DR WPI; 2003-332730/31.
 XX
 PT New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 PS Claim 57; SEQ ID NO 277; 169pp; English.
 XX
 CC The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 18 AA;
 Query Match 7.4%; Score 18; DB 7; Length 18;
 Best Local Similarity 100.0%; Pred. No. 8.7e-11;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 210 LNPFMRKPIPTVZAG 227
 Db 1 LNPFMRKPIPTVZAG 18
 RESULT 72
 ADCL17649
 ID ADCL17649 standard; peptide; 18 AA.
 AC ADCL17649;
 XX
 XX 18-DEC-2003 (first entry)
 XX
 XX Type IV collagen NC1 domain related peptide SEQ ID NO:254.
 XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO2003012122-A2.
 XX
 PD 13-FEB-2003.
 XX
 PD 26-JUL-2002; 2002WO-US023763.
 XX
 PR 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX
 PI Sundaramoorthy M, Hudson B;
 XX
 DR WPI; 2003-332730/31.
 XX
 PT New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 PS Claim 57; SEQ ID NO 254; 169pp; English.
 XX
 CC The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue

CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.

XX Sequence 18 AA;

Query Match 7.4%; Score 18; DB 7; Length 18;

Best Local Similarity 100.0%; Pred. No. 8.7e-11;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTMPFLFCNVNDVGNFA 88

Db 1 FTMPFLFCNVNDVGNFA 18

RESULT 73

AA118657

ID AA118657 standard; protein; 46 AA.

XX AA118657;

AC AA118657;

DT 12-OCT-2001 (first entry)

XX Peptide #5091 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX Cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000670.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234487P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
 CC gene expression in human cervical epithelial cells.

XX Claim 27; SEQ ID NO 23483; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs: see AA118657-A128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 46 AA;

Query Match 7.0%; Score 17; DB 4; Length 46;

Best Local Similarity 100.0%; Pred. No. 2.2e-09;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100

Db 10 VCNFASRNDYSYWLSTP 26

RESULT 74

ABG40558

ID ABG40558 standard; peptide; 46 AA.

XX ABG40558;

XX 19-AUG-2002 (first entry)

XX Human peptide encoded by genome-derived single exon probe SEQ ID 30223.

XX Human; single exon probe; asthma; lung cancer; COPD; ILD;

XX chronic obstructive pulmonary disease; interstitial lung disease;

XX familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhage;

XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX primary ciliary dyskinesia; pulmonary hypertension;

XX hyaline membrane disease.

XX Homo sapiens.

XX WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US000665.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234487P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to
 CC measure gene expression in human lung samples.

XX Claim 27; SEQ ID NO 30223; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC ; the novel set of probes which hybridize at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: the sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 46 AA;

Query Match 7.0%; Score 17; DB 5; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
| | | | | | | | | | | | | | | | | | | | | |
DB 10 VCNFASRNDYSYWLSTP 26

RESULT 75

AAP93524
ID AAP93524 standard; protein; 229 AA.

XX AC AAP93524;

XX DT 25-MAR-2003 (revised)

XX DT 03-OCT-2002 (revised)

XX DT 04-JUN-1990 (first entry)

XX Complete sequence of the alpha-1-NC1 domain of type IV collagen.

XX Alpha-1-NC1 domain; type IV collagen; cell adhesion; heparin;
XX aortic endothelial cells; metastatic carcinoma M4 cells; rat fibroblasts;
XX MM fibrosarcoma cells; C6 glioma cell; A431 breast carcinoma cells;
XX wound healing; implant acceptance.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Peptide 17..27

XX FT /note= "TS-3"

XX FT Peptide 49..60

XX FT /note= "TS-2"

XX FT Peptide 201..216

XX FT /note= "TS-1"

XX PN W08903392-A.

XX PD 20-APR-1989.

XX PF 20-AUG-1988; 88WO-US003023.

XX PR 08-OCT-1987; 87US-00106858.

XX PA (MINU) MINNESOTA UNIVERSITY.

XX PI Tailbary EC;

XX DR WPI; 1989-130015/17.

XX Polypeptide(s) with type IV collagen activity - used to promote wound

PT healing, implant acceptance and cellular attachment and inhibit malignant
PT cells.

XX PS Fig 2; page 1/12; 40pp; English.

XX The peptides in the features table are claimed (Claim 1, p. 22). They
CC were synthesised using the Merrifield solid phase method. Binding assays
CC were carried out using peptides TS-1, TS-2 and TS-3. TS-1 promotes
CC adhesions of aortic endothelial cells, metastatic carcinoma M4 cells,
CC normal rat fibroblasts, MM fibrosarcoma cells, C6 glioma cells and A431
CC breast carcinoma cells. TS-2 binds to type IV collagen, to heparin and
CC promotes adhesion of the above cells. Peptides TS-1, TS-2 and TS-3 may be
CC used to promote wound healing and implant acceptance, promote cellular
CC attachment to culture substrata or inhibit the metastasis of malignant
CC cells. They may be used to coat a prosthetic device. (Updated on 03-OCT-
CC 2002 to add missing OS field.) (Updated on 25-MAR-2003 to correct PF
CC field.) (Updated on 25-MAR-2003 to correct PI field.)

XX SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.2e-09;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100

DB 70 VCNFASRNDYSYWLSTP 86

RESULT 76

AAY67943

ID AAY67943 standard; protein; 229 AA.

XX AC AAY67943;

XX DT 03-APR-2000 (first entry)

XX Human type IV collagen alpha 1 chain protein sequence SEQ ID NO:2.

XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
XX benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
XX ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;
XX myocardial angiogenesis; plaque neovascularisation; angiodioma;
XX atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
XX contraception; obesity.

XX OS Homo sapiens.

XX PN W09965940-A1.

XX PD 23-DEC-1999.

XX PF 17-JUN-1999; 99WO-US013737.

XX PR 17-JUN-1998; 98US-0089689P.

XX PR 25-MAR-1999; 99US-0126175P.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2000-097708/08.

XX DR N-PSDB; AA257159.

XX Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or 3
PT chain of Type IV collagen used in, e.g. treatment of benign tumors and
PT rheumatoid arthritis.

XX PS Example 1; Fig 1B; 117pp; English.

XX The present sequence represents the human type IV collagen alpha 1 chain.
CC The present invention describes an isolated protein chosen from the NC1
CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a

CC fragment, analogue, derivative or mutant, which has anti-angiogenic
 CC properties. The anti-angiogenic proteins, multimers and chimeras are
 CC useful for inhibiting angiogenic activity in mammalian tissue, especially
 CC for treating diseases chosen from angiogenesis-dependent cancers, benign
 CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular
 CC angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiodysplasia, wound granulation, intestinal adhesions, atherosclerosis,
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori
 CC ulcers, dialysis graft vascular access stenosis, contraception and
 CC obesity. The compositions can be used to inhibit a disease characterised
 CC by angiogenic activity, in conjunction with radiation therapy,
 CC chemotherapy or immunotherapy

XX Sequence 229 AA;

Query Match 7.0%; Score 17; DB 3; Length 229;
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
 |||||
 Db 70 VCNFASRNDYSYWLSTP 86

RESULT 77

AAU75587

ID AAU75587 standard; protein; 229 AA.

AC AAU75587;

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 1 chain.

XX Human; type IV collagen alpha 1 chain; cytostatic; antiangiogenic;
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
 KW Tumstatin; angiogenesis; tumour.

OS Homo sapiens.

PN WO200151523-A2.

PD 19-JUL-2001.

PP 08-JAN-2001; 2001WO-US000565.

PR 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PI Kalluri R;

XX WPI; 2002-188037/24.

DR N-PSDB; ABK15359.

PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
 PT treating disorders involving angiogenesis.

PS Example 1; Fig 1B; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
 CC domain, having one or more of the characteristics selected from: (a) the
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
 CC proliferation of endothelial cells; and (c) the ability to cause
 CC apoptosis of endothelial cells. Also described are the following: (1) use
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
 CC analogue or allelic variant in the preparation of a medicament for
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
 CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the
 CC endothelial cell apoptosis is mediated by one or more endothelial cell
 CC integrins or one or more endothelial cell integrin subunits; (2) use of
 CC an antibody or peptide that specifically binds the alpha1, alpha2,
 CC alpha3, alpha4, alpha5, alpha6, beta1 or beta3 subunit of integrin in the
 CC preparation of a medicament for inhibiting angiogenesis or cell
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation
 CC of a medicament for treating a proliferative disease in a vertebrate,
 CC where the disease is characterised by angiogenesis that is mediated by
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
 CC the presence of a medicament for promoting angiogenesis in a tissue; and
 CC (5) use of integrins in the preparation of a medicament for promoting or
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
 CC or allelic variants are useful in the preparation of a medicament for
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where
 CC the angiogenesis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits; or by promoting or
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial
 CC cell apoptosis is mediated by one or more endothelial cell integrins or
 CC one or more endothelial cell integrin subunits. The medicament is useful
 CC in inhibiting tumour growth and for the regression of an established
 CC tumour. The present sequence represents the amino acid sequence of human
 CC type IV collagen alpha 1 chain

XX Sequence 229 AA;

Query Match 7.0%; Score 17; DB 5; Length 229;
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100

|||||
 Db 70 VCNFASRNDYSYWLSTP 86

RESULT 78

ADA20217

ID ADA20217 standard; protein; 229 AA.

AC ADA20217;

XX 20-NOV-2003 (first entry)

DE Human type IV collagen alpha 1 chain partial protein sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; alpha 1 chain; arresten; human.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20216.

XX

PT New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX
 XX
 PS Claim 101; Fig 1; 240pp; English.
 XX
 CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumor growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumor
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the partial amino acid sequence of the alpha 1 chain of human
 CC type IV collagen. The "arresten" peptide of the invention was derived
 CC from this protein.
 XX
 SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 6; Length 229;
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
 DB 70 VCNFASRNDYSYWLSTP 86

RESULT 79

ID ADC17695 standard; protein; 229 AA.
 XX
 AC ADC17695;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human type IV collagen alpha 1 chain protein SEQ ID NO:302.
 XX
 KW crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumor metastasis inhibitor; tumor growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glioma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO2003012122-A2.
 XX
 PD 13-FEB-2003.
 XX
 PF 26-JUL-2002; 2002WO-US023763.
 XX
 PR 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-0351289P.
 PR 02-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUSD/) HUDSON B.

XX Sundaramoorthy M, Hudson B;
 XX WPI; 2003-332730/31.
 XX
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glioma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 PS Disclosure; SEQ ID NO 302; 168pp; English.
 XX
 CC The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antipsoriatic, antianemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glioma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents an amino acid sequence which is used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 7; Length 229;
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
 DB 70 VCNFASRNDYSYWLSTP 86

RESULT 80

ID ADC17699
 XX ADC17699 standard; protein; 229 AA.
 AC ADC17699;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human type IV collagen alpha 5 chain protein SEQ ID NO:306.
 XX
 KW crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumor metastasis inhibitor; tumor growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glioma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO2003012122-A2.
 XX

PD 13-FEB-2003.
 XX 26-JUL-2002; 2002WO-US023763.
 XX 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-03511289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND// SUNDARAMOORTHY M.
 PA (HUDS// HUDSON B.
 XX Sundaramoorthy M, Hudson B;
 XX WPI; 2003-332730/31.
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 XX Disclosure; SEQ ID NO 306; 168pp; English.
 XX
 XX The present invention describes a crystallised NCI domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (5) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NCI
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
 CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and
 CC anticancer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents an amino acid sequence which is used in the exemplification of
 XX the present invention.
 XX Sequence 229 AA;
 SQ
 Query Match 7.0%; Score 17; DB 7; Length 229;
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 84 VCNFASRNDYSYWLSTP 100
 DB 70 VCNFASRNDYSYWLSTP 86
 |||||
 RESULT 81
 AAY31991
 ID AAY31991 standard; protein; 260 AA.
 XX AAY31991;
 AC AAY31991;
 XX
 XX 05-JAN-2000 (first entry)
 DT Type IV collagen NCI domain alpha-1 monomer.
 XX
 DE Type IV collagen; NCI domain; non-collagenous domain; human;
 XX
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
 KW rheumatoid arthritis; retinal neovascularization;
 KW

KW chorioidal neovascularization; macular degeneration;
 KW corneal neovascularization; retinopathy of prematurity;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidemic keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; sogrens; acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW ulcer; herpes simplex infection; Herpes zoster infection;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;
 KW sarcoïd; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu; AIDS;
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
 KW pemphigoid.
 XX
 XX Homo sapiens.
 OS Synthetic.
 XX
 XX Location/Qualifiers
 XX Key
 XX Peptide 1..17 /note= "BM40 signal peptide"
 XX Protein 18..260 /note= "mature protein"
 XX Peptide 18..25 /note= "affinity tag"
 XX Protein 26..260 /note= "NCI alpha-1 monomer"
 XX
 XX WO9949885-A2.
 XX
 XX 07-OCT-1999.
 PD
 XX 26-MAR-1999; 99WO-US006445.
 XX
 XX 27-MAR-1998; 98US-0079783P.
 PR 29-OCT-1998; 98US-0106170P.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 XX Hudson BG, Sarrae MP;
 PI
 XX WPI; 1999-601297/51.
 DR N-PSDB; AA220089.
 DR
 XX Inhibition of angiogenesis with non-collagenous alpha chain monomer
 PT useful for treating e.g. tumor growth or metastasis, neovascularisation,
 PT etc.
 XX
 XX Disclosure; Fig 17a; 56pp; English.
 PS
 XX This sequence represents a recombinant type IV collagen non-collagenous
 CC (NCI) domain alpha-1 polypeptide composed of a BM40 signal sequence
 CC (which is cleaved from the mature protein) to facilitate protein
 CC secretion, and a mature protein comprising an affinity tag (facilitates
 CC purification and identification of the material) and the alpha-1 chain
 CC monomer. The invention provides methods and kits for inhibiting
 CC angiogenesis, tumour growth and metastasis, and endothelial cell
 CC interaction with the extracellular matrix, each method comprising
 CC contacting the tumour or animal tissue with 1 or more isolated type IV
 CC collagen NCI alpha chain monomer(s) selected from the group consisting of
 CC alpha-1, alpha-2, alpha-3 and alpha-6 NCI chain monomers (see AAY31991-
 CC 96). The monomers can be produced via recombinant protein expression. The
 CC polynucleotides and polypeptides are used to treat an angiogenesis
 CC mediated disorder or condition, especially selected from solid and blood-
 CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
 CC neovascularization, choroidal neovascularization, macular degeneration,

CC corneal neovascularization, retinopathy of prematurity, corneal graft
 CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
 CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, sogrens,
 CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
 CC simplex infections, herpes zoster infections, protozoan infections,
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
 CC keratolysis, trauma, systemic lupus, pterygoiditis, Wegener's
 CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,
 CC sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Paget's disease,
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets
 CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
 CC complications, abnormal proliferation of fibrovascular tissue
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)
 XX
 SQ Sequence 260 AA;

Query Match 7.0%; Score 17; DB 2; Length 260;
 Best Local Similarity 100.0%; Pred. No. 1e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
 |||||
 DB 101 VCNFASRNDYSYWLSTP 117

RESULT 82
 AAY97553
 ID AAY97553 standard; protein; 260 AA.
 XX
 AC AAY97553;

DT 12-FEB-2001 (first entry)

XX Human alpha(IV)NC1 protein sequence.

XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
 KW tumor growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
 KW chronic inflammation; psoriasis; therapy; alpha(IV)NC1.

OS Homo sapiens.

XX WO200059532-A1.

XX 12-OCT-2000.

PF 31-MAR-2000; 2000WO-US008678.

XX 01-APR-1999; 99US-0127391P.

XX (BIOS-) BIOSTRATUM INC.

XX Brooks P, Hudson B;

DR WPI; 2000-664962/64.

XX N-PSDB; AAA90991.

PT Use of antagonists of specific integrin receptors for inhibiting
 PT angiogenesis, tumor growth or metastases, or endothelial cell
 PT interactions with the extracellular matrix.

PS Disclosure; Fig 17a; 78pp; English.

XX

CC This sequence is a human type IV collagen alpha chain monomer, designated
 CC alpha(IV)NC1. The invention relates to a method for inhibiting
 CC angiogenesis, tumor growth or metastases, or endothelial cell
 CC interactions with the extracellular matrix, comprising contacting the
 CC cells or tissue with a polypeptide composition containing antagonists of
 CC specific integrin receptors. The methods and the antagonists are useful
 CC for inhibiting angiogenesis, tumor growth or metastases, or endothelial
 CC cell interaction with the extracellular matrix. The antagonists are also
 CC useful for treating diseases and conditions with accompanying undesired
 CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
 CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
 CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
 CC treating non-tumorigenic diseases and conditions with accompanying
 CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
 CC retinal neovascularisation, choroidal neovascularisation, muscular
 CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
 CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
 CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
 CC complications, chronic inflammation or psoriasis
 XX
 SQ Sequence 260 AA;

Query Match 7.0%; Score 17; DB 3; Length 260;
 Best Local Similarity 100.0%; Pred. No. 1e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
 |||||
 DB 101 VCNFASRNDYSYWLSTP 117

RESULT 83
 AAY31995
 ID AAY31995 standard; protein; 264 AA.
 XX
 AC AAY31995;

DT 05-JAN-2000 (first entry)

XX Type IV collagen NC1 domain alpha-5 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;
 KW angiogenesis; tumor; metastasis; therapy; diabetic retinopathy;
 KW rheumatoid arthritis; retinal neovascularization;
 KW choroidal neovascularization; macular degeneration;
 KW corneal neovascularization; retinopathy of prematurity;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidemic keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; sogrens; acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW ulcer; herpes simplex infection; Herpes zoster infection;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; pterygoiditis; Wegener's sarcooidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;
 KW sarcoid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu; AIDS;
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
 KW pemphigoid.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..17

FT Protein /note= "BM40 signal peptide"

18..264


```
RESULT 85
AAB54044
ID AAB54044 standard; protein; 309 AA.
XX AC AAB54044;
XX DT 09-MAR-2001 (first entry)
XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:496.
XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
XX KW diagnosis; identification; cytostatic; neuroprotective; nontropic;
XX KW immunomodulatory; relaxant; contraceptive; gynaecological;
XX KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
XX KW linkage analysis; tissue identification; tissue typing; forensic; neural;
XX KW immune system; muscular; reproductive; gastrointestinal; pulmonary;
XX KW cardiovascular; renal; proliferative.
XX OS Homo sapiens.
XX PN WO200055320-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005989.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2000-579444/54.
XX DR N-PSDB; AAC98809.
XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
XX PT treating, or ameliorating a medical condition, particular pancreatic
XX PT cancer, or for use in assays for diagnosing a pathological condition.
XX PS Claim 11; Page 934-935; 1379pp; English.
XX CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX CC proteins, called pancreatic cancer antigens, given in AAB54008 to
XX CC AAB54466. The human pancreatic cancer antigens have cytostatic,
XX CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
XX CC gynaecological, cardiant and antiinflammatory activities, and can be used
XX CC in gene therapy. The polynucleotide and proteins can be used for
XX CC preventing, treating, or ameliorating a medical condition or in assays
XX CC for diagnosing a pathological condition or a susceptibility to one in a
XX CC subject. Binding partners to the proteins and the activity of the
XX CC proteins can be identified. The pancreatic cancer antigens can be used to
XX CC detect, treat or prevent pancreatic disorders, especially cancer.
XX CC Agonists and antagonists to the antigens can be screened for. The
XX CC pancreatic cancer antigen polynucleotides can be used to design nucleic
XX CC acid hybridisation probes that can be used in chromosome mapping, linkage
XX CC analysis, tissue identification and/or typing and a variety of forensic
XX CC and diagnostic methods. The proteins can be used to generate antibodies
XX CC which are used to purify, detect and target the polypeptides, including
XX CC both in vivo and in vitro diagnostic and therapeutic methods. The
XX CC proteins can be used to treat or prevent neural, immune system, muscular,
XX CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX CC sequences used in the exemplification of the present invention
XX SQ Sequence 309 AA;
Query Match 7.0%; Score 17; DB 3; Length 309;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100

RESULT 86
AAB58169
ID AAB58169 standard; protein; 406 AA.
XX AC AAB58169;
XX DT 14-MAR-2001 (first entry)
XX DE Lung cancer associated polypeptide sequence SEQ ID 507.
XX KW Human; lung cancer associated protein; neuroprotective; cytostatic;
XX KW cardioactive; immunomodulatory; muscular active; vulnerary;
XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;
XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX KW proliferative disorder; wound healing; infectious disease.
XX OS Homo sapiens.
XX PN WO200055180-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005918.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX DR WPI; 2000-587514/55.
XX DR N-PSDB; AAF18045.
XX PT Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer.
XX PS Claim 11; Page 996-998; 1425pp; English.
XX CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective, cytostatic; cardioactive;
XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the protein
XX CC or polynucleotide sequences. The lung cancer associated polynucleotide
XX CC sequences may be used for detection of lung cancer, chromosome
XX CC identification, as chromosome markers, and for numerous other diagnostic
XX CC or research purposes. The proteins may be used to treat disorders such as
XX CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
XX CC cardiovascular, renal, and proliferative disorders. The proteins may also
XX CC be used in the treatment of wounds and infectious diseases.
XX CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
XX CC used in the course of the invention for the identification and
XX CC characterisation of the polynucleotide and protein sequences
XX SQ Sequence 406 AA;
Query Match 7.0%; Score 17; DB 3; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
Db 247 VCNFASRNDYSYWLSTP 263
```

```

RESULT 87
AAR23873
ID AAR23873 standard; protein; 772 AA.
XX
XX
AC AAR23873;
XX
XX
DT 25-NOV-1992 (first entry)
XX
XX
DE Human alpha 5 (IV) of type IV collagen.
XX
XX
KW Mutations; Alport's syndrome; basement membranes; diabetes mellitus.
XX
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 43..47
FT /note= "interruption in Gly-X-Y sequence"
FT Misc-difference 159..160
FT /note= "interruption in Gly-X-Y sequence"
FT Misc-difference 275..277
FT /note= "interruption in Gly-X-Y sequence"
FT Misc-difference 334..336
FT /note= "interruption in Gly-X-Y sequence"
FT Misc-difference 456..458
FT /note= "interruption in Gly-X-Y sequence"
XX
XX
US5114840-A.
XX
XX
PD 19-MAY-1992.
XX
XX
PF 07-JUL-1989; 89US-00377238.
XX
XX
PR 07-JUL-1989; 89US-00377238.
XX
XX
PA (TRYG/) TRYGGVASON K.
XX
XX
PI Tryggvason K, Hostikka SL;
XX
XX
WPI; 1992-192174/23.
XX
XX
N-PSDB; AAR24551.
XX
XX
Isolation of DNA encoding alpha-5(IV) polypeptide of type IV collagen - to
detect mutations in genes for alpha-5(IV) chain which produce genetic or
acquired basement membrane disorders e.g. Alport's syndrome.
XX
XX
PS Disclosure; Fig 2; 14pp; English.
XX
XX
The sequence is that of the alpha 5(IV) polypeptide chain of human type
IV collagen, the major component of basement membranes. The protein
contains the Gly-X-Y repeat coding sequence typical for collagenous
proteins at one end and a typical NC-domain coding sequence at the other
end. The sequence can be used to detect mutations in individual genes
specific for this chain which can, directly or indirectly, produce
several human diseases. It can also be used to determine genetic, e.g.
Alport's syndrome, or acquired e.g. diabetes mellitus, disorders of the
basement membrane, and as probes or antibodies against these nucleotide
sequences. Gene fragments generated through amplifications from human
genomic or cloned DNA can also be used for detection and analysis of
genes
XX
XX
Sequence 772 AA;
Query Match 7.0%; Score 17; DB 2; Length 772;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 84 VCNFASRNDYSYWLSTP 100
Db 613 VCNFASRNDYSYWLSTP 629
RESULT 88
AAW09643
ID AAW09643 standard; protein; 772 AA.
XX
XX
AC AAW09643;
XX
XX
DT 25-MAR-2003 (revised)
DT 16-JUN-1997 (first entry)
XX
XX
DE Human type IV collagen alpha-5.
XX
XX
KW Collagen alpha5(IV); basement membrane; Alport's syndrome; nephritis;
kidney; renal failure; antibody; diagnosis; COL4A5 gene; X chromosome.
XX
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT /label= Collagenous domain
FT /note= "collagenous" domain contains Gly-X-Y tripeptide
repeats, interrupted at positions 43-47, 159-160, 275-
276, 334-335, 456-459"
FT Domain 544..772
FT /label= Non-collagenous_domain
FT Peptide 742..751
FT /label= Immunogenic peptide
FT /note= "peptide used to raise diagnostic antibodies
(Claim 1)"
XX
XX
US5593900-A.
XX
XX
PD 14-JAN-1997.
XX
XX
PF 11-OCT-1994; 94US-00321084.
XX
XX
PR 07-JUL-1989; 89US-00377238.
XX
XX
PR 20-DEC-1990; 90US-00630563.
XX
XX
(TRYG/) TRYGGVASON K.
XX
XX
(HOST/) HOSTIKKA S L.
XX
XX
(HOYH/) HOYHTYA M.
XX
XX
Hostikka SL, Tryggvason K, Hoyhtya M;
XX
XX
WPI; 1997-099481/09.
XX
XX
N-PSDB; AAT47812.
XX
XX
New antibodies specific for human type IV collagen alpha5 chain - used to
detect absence of this chain in patients with renal failure.
XX
XX
PS Disclosure; Fig 2A-2B; 12pp; English.
XX
XX
The amino acid sequence of a portion (AAW09643) of the previously unknown
human type IV collagen chain, alpha5(IV), was deduced from cDNA clones
(see also AAT47812) obt'd using probes based on conserved sequences of
human alpha1(IV) and alpha2(IV) collagen chains and of the Drosophila
alpha(IV) chain. It includes a complete non-collagenous domain that shows
83% identity with that of alpha1(IV) and 63% with that of the alpha2(IV)
chain. Mutations in the alpha5(IV) gene (COL4A5) are associated with
Alport's syndrome. Antibodies raised against a peptide (see also
AAW09644) specific to alpha5(IV) can be used in the diagnosis of basement
membrane disorders such as Alport's syndrome. (Updated on 25-MAR-2003 to
correct PF field.)
XX
XX
Sequence 772 AA;
Query Match 7.0%; Score 17; DB 2; Length 772;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 84 VCNFASRNDYSYWLSTP 100
Db 613 VCNFASRNDYSYWLSTP 629

```

Qy 84 VCNFASRNDYSYWLSTP 100

AC ABB57334;
XX
XX
XX 07-MAR-2002 (first entry)
XX
XX Mouse ischaemic condition related protein sequence SEQ ID NO:933.
DE
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic ischaemia; ischaemic condition; ischaemic disease.
KW
XX
XX Mus musculus.
XX
XX WO200188188-A2.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-JP004192.
PF
XX
XX 18-MAY-2000; 2000JP-00145977.
PR
XX
XX (UYN1-) UNIV NIHOON SCHOOL JURIDICAL PERSON.
PA
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
XX WPI; 2002-034733/04.
PI
XX
XX N-PSDB; ABI99819.
DR
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or by
PT determining the expression profile of a gene group comprising these
PT genes.
XX
XX Claim 2; Page 2352-2359; 2690pp; English.
PS
XX
XX The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI9912, encoding the
CC protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The expression
CC levels or expression profiles produced by these genes are used as an
CC indicator when screening for ischaemic condition-improving drugs or
CC therapeutics for ischaemic diseases. ABI9913 and ABI9914 represent PCR
CC primers for a mouse ischaemic condition related sequence, which are used
CC in the exemplification of the present invention
XX
XX SQ Sequence 1669 AA;
Query Match 7.0%; Score 17; DB 5; Length 1669;
Best Local Similarity 100.0%; Pred. No. 5.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
DB 1510 VCNFASRNDYSYWLSTP 1526
RESULT 92
ABU54467
ID ABU54467 standard; protein; 1669 AA.
XX
XX ABU54467;
AC
XX
XX 12-MAR-2003 (first entry)
DT
XX
XX Human tumour endothelial marker TEM 31.
DE
XX
XX Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;
KW Tumour endothelial marker; normal endothelial marker; PEM;
KW pan-endothelial marker; polycystic kidney disease; psoriasis;
KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;
KW

KW neoangiogenesis; immune response; cytostatic; antidiabetic;
KW ophthalmological; antirheumatic; antiarthritic; antipsoriatic.
XX
XX Homo sapiens.
OS
XX
XX WO200283874-A2.
PN
XX
XX 24-OCT-2002.
PD
XX
XX 10-APR-2002; 2002WO-US008253.
PF
XX
XX 11-APR-2001; 2001US-0282850P.
PR
XX
XX 06-FEB-2002; 2002US-0354262P.
PR
XX
XX (UYJO) UNIV JOHNS HOPKINS.
PA
XX
XX Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;
PI
XX
XX WPI; 2003-093016/08.
DR
XX
XX N-PSDB; ABX72039.
DR
XX
XX New purified human transmembrane protein, designated as tumor endothelial
PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors, or
PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or
PT psoriasis.
XX
XX PS Disclosure; Page 272-275; 374pp; English.
XX
XX The present invention relates to a novel method for the isolation of
CC endothelial cells (ECs), and the identification of genes expressed in
CC normal and tumour ECs. Tumour endothelial marker (TEM), normal
CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are
CC identified in human ECs. The human EC marker proteins and the
CC polynucleotide sequences encoding them are useful for detecting,
CC diagnosing or treating tumors as well as polycystic kidney disease,
CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also
CC useful for inhibiting neoangiogenesis or tumour angiogenesis, for
CC inducing an immune response to tumour endothelial cells in a patient, or
CC for identifying candidate drugs for treating tumors. The present
CC sequence represents a human TEM or NEM protein of the invention
XX
XX SQ Sequence 1669 AA;
Query Match 7.0%; Score 17; DB 6; Length 1669;
Best Local Similarity 100.0%; Pred. No. 5.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
DB 1510 VCNFASRNDYSYWLSTP 1526
RESULT 93
AAM39077
ID AAM39077 standard; protein; 1672 AA.
XX
XX AAM39077;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 2222.
DE
XX
XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX OS Homo sapiens.
XX
XX WO200153312-A1.
PN
XX

```
PD 26-JUL-2001.
XX
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX
XX 21-JAN-2000; 2000US-00488725.
XX
XX 25-APR-2000; 2000US-00552317.
XX
XX 20-JUN-2000; 2000US-00598042.
XX
XX 19-JUL-2000; 2000US-00620312.
XX
XX 03-AUG-2000; 2000US-00653450.
XX
XX 14-SEP-2000; 2000US-00662191.
XX
XX 19-OCT-2000; 2000US-00693036.
XX
XX 29-NOV-2000; 2000US-00727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX
XX N-PSDB; AA158233.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2222; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
XX encoded polypeptides (AA158642-AA162213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX
XX Sequence 1672 AA;
XX
XX Query Match 7.0%; Score 17; DB 4; Length 1672;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-08;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 84 VCNFASRNDYSYWLSTP 100
XX |||||
XX Db 1513 VCNFASRNDYSYWLSTP 1529
XX
XX RESULT 94
XX ABG04839
XX ID ABG04839 standard; protein; 1685 AA.
XX
XX AC ABG04839;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #4830.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX DR N-PSDB; AAS69026.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 35198; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activities in
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX patent did not appear in the invention. Note: The sequence data for this
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1685 AA;
XX
XX Query Match 7.0%; Score 17; DB 4; Length 1685;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-08;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 84 VCNFASRNDYSYWLSTP 100
XX |||||
XX Db 1526 VCNFASRNDYSYWLSTP 1542
XX
XX RESULT 95
XX ABG15619
XX ID ABG15619 standard; protein; 1693 AA.
XX
XX AC ABG15619;
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #15610.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
```

```
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanan RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS79806.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 45978; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SEQ Sequence 1693 AA;
XX
XX Query Match 7.0%; Score 17; DB 4; Length 1693;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-08;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 84 VCNFASRNDXYWLTSTP 100
XX | | | | | | | | | | | | | | | | | |
XX Db 1534 VCNFASRNDXYWLTSTP 1550
XX
XX RESULT 96
XX ADC17470
XX ID ADC17470 standard; peptide; 16 AA.
XX
XX AC ADC17470;
XX
XX AC ADC17470;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:74.
XX
XX KW crystallised NC1 domain hexamer of type IV collagen;
XX KW angiogenesis inhibitor; angiogenesis-mediated disease;
XX KW tumour metastasis inhibitor; tumour growth inhibitor;
XX KW endothelial cell interaction inhibitor;
XX KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
XX KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX KW endothelial cell adhesion inhibitor;
XX KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
XX KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX KW blood-borne tumour.
XX
XX OS Synthetic.
```

```
OS Homo sapiens.
XX WO2003012122-A2.
XX
XX PD 13-FEB-2003.
XX
XX PF 26-JUL-2002; 2002WO-US023763.
XX
XX PR 27-JUL-2001; 2001US-0308523P.
XX PR 29-OCT-2001; 2001US-0351289P.
XX PR 22-MAR-2002; 2002US-0366854P.
XX PR 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT basal lamina membrane formation in cell or tissue development.
XX
XX Disclosure; SEQ ID NO 74; 169pp; English.
XX
XX The present invention describes a crystallised NC1 domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (6) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NC1
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
CC antipsoriatic, antianemic, ophthalmological, antiarteriosclerotic and
CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
CC invention.
XX
XX SEQ Sequence 16 AA;
XX
XX Query Match 6.6%; Score 16; DB 7; Length 16;
XX Best Local Similarity 100.0%; Pred. No. 8.8e-09;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 50 VQGNQRAHGQDLGTLG 65
XX | | | | | | | | | | | | | | | | | |
XX Db 1 VQGNQRAHGQDLGTLG 16
XX
XX RESULT 97
XX AAE09491
XX ID AAE09491 standard; peptide; 15 AA.
XX
XX AC AAE09491;
XX
XX DT 19-NOV-2001 (first entry)
XX
XX DE Human C2 alpha-3 peptide to construct alpha/alpha3 (IV)NC1 protein.
```


XX The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is human alpha chain peptide used for constructing human
CC alpha1/alpha3 (IV) NCI fusion protein
XX
SQ Sequence 15 AA;

Query Match 6.1%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 8.8e-08; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 TDIPPCPHGWISLWK 153
Db 1 TDIPPCPHGWISLWK 15
|||||

RESULT 100
AAE09495
ID AAE09495 standard; peptide; 15 AA.
XX AC AAE09495;
XX 19-NOV-2001 (first entry)
XX Human C4 alpha-3 peptide to construct alpha1/alpha3 (IV) NCI protein.
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
XX Goodpasture syndrome; C4 alpha-3 peptide.
XX Homo sapiens.
XX US6277558-B1.
XX 21-AUG-2001.
XX 12-NOV-1999; 99US-00439897.
XX 30-NOV-1990; 90US-00621091.
XX 07-MAR-1995; 95US-00399889.
XX 07-OCT-1998; 98US-00167364.
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX Hudson BG;
XX WPI; 2001-540401/60.
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
XX Goodpasture antibodies from bodily fluid/tissue from patient or for
XX treating Goodpasture syndrome by contacting bodily fluid or tissue with
XX the polypeptide.
XX Example 19; Fig 12; 46pp; English.
XX The invention relates to a method for detecting Goodpasture antibodies
XX from a bodily fluid or tissue of a patient. The method comprises
XX contacting the bodily fluid or tissue with alpha-3 chain type (IV)
XX collagen polypeptide that contains a conformational epitope for the
XX Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
XX detecting Goodpasture antibodies from a bodily fluid or tissue from a
XX patient, and for treating Goodpasture syndrome in a patient. The present
XX sequence is human alpha chain peptide used for constructing human
XX alpha1/alpha3 (IV) NCI fusion protein
XX
SQ Sequence 15 AA;

Query Match 6.1%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 8.8e-08; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 8.8e-08; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EKIISRCQVCMKGRH 244
Db 1 EKIISRCQVCMKGRH 15
|||||

RESULT 101
ADC17449
ID ADC17449 standard; peptide; 15 AA.
XX AC ADC17449;
XX 18-DEC-2003 (first entry)
XX Type IV collagen NCI domain related peptide SEQ ID NO:53.
XX crystallised NCI domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
XX tumour metastasis inhibitor; tumour growth inhibitor;
XX endothelial cell interaction inhibitor;
XX basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
XX antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX endothelial cell adhesion inhibitor;
XX endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
XX ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX blood-borne tumour.
XX Synthetic.
XX Homo sapiens.
XX WO2003012122-A2.
XX 13-FEB-2003.
XX 26-JUL-2002; 2002WO-US023763.
XX 27-JUL-2001; 2001US-0308523P.
XX 29-OCT-2001; 2001US-0351289P.
XX 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUSD/) HUDSON B.
XX Sundaramoorthy M, Hudson B;
XX WPI; 2003-332730/31.
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX Disclosure; SEQ ID NO 53; 168pp; English.
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (6) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
XX antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX antiulcer activities, and can be used as an inhibitor of angiogenesis,
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition

CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 15 AA;

Query Match 6.1%; Score 15; DB 7; Length 15;
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SLNPFMRKPIPT 223
 DB 1 SLNPFMRKPIPT 15
 |||||

RESULT 102

ADCL17586
 ID ADCL17586 standard; peptide; 15 AA.

XX AC ADCL17586;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:191.

XX KW crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

XX OS Synthetic.

OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX PA (UNIV) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUSD/) HUDSON B.

XX PI Sundaramoorthy M, Hudson B;

XX DR WPI; 2003-332730/31.

XX PT New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.

XX PS Disclosure; SEQ ID NO 191; 168pp; English.

XX CC The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)

CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (5) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (1) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 15 AA;

Query Match 6.1%; Score 15; DB 7; Length 15;
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 FRKPIPTVKAGELE 230
 DB 1 FRKPIPTVKAGELE 15
 |||||

RESULT 103

ADCL17607

ID ADCL17607 standard; peptide; 15 AA.

XX AC ADCL17607;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:212.

XX KW crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

XX OS Synthetic.

OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX PA (UNIV) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUSD/) HUDSON B.

XX

KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
KW anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.
XX
OS Synthetic.
OS Homo sapiens.
XX WO2003012122-A2.
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
XX 29-OCT-2001; 2001US-0351289P.
XX 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUSD/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Disclosure; SEQ ID NO 137; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
XX antipsoriatic, anti-anaemic, ophthalmological, antiarteriosclerotic and
XX antiulcer activities, and can be used as an inhibitor of angiogenesis,
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 15 AA;
XX
XX Query Match 6.1%; Score 15; DB 7; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 8.Be-08;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 100 PALMPNNWAPITGRA 114
XX |||||
XX DB 1 PALMPNNWAPITGRA 15

RESULT 106
ADCL17490
ID ADCL17490 standard; peptide; 15 AA.
XX
AC ADCL17490;
XX
XX 18-DEC-2003 (first entry)
XX
XX Type IV collagen NCI domain related peptide SEQ ID NO:94.
XX
XX crystallised NCI domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
XX tumour metastasis inhibitor; tumour growth inhibitor;
XX endothelial cell interaction inhibitor;
XX basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
XX anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX endothelial cell adhesion inhibitor;
XX endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
XX ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX blood-borne tumour.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX WO2003012122-A2.
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
XX 29-OCT-2001; 2001US-0351289P.
XX 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUSD/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Disclosure; SEQ ID NO 94; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
XX antipsoriatic, anti-anaemic, ophthalmological, antiarteriosclerotic and
XX antiulcer activities, and can be used as an inhibitor of angiogenesis,
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 15 AA;
XX
XX Query Match 6.1%; Score 15; DB 7; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 8.Be-08;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 100 PALMPNNWAPITGRA 114
XX |||||
XX DB 1 PALMPNNWAPITGRA 15

PS Disclosure; SEQ ID NO 28; 169pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (5) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NC1
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
CC antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
CC antitumor activities, and can be used as an inhibitor of angiogenesis,
CC tumor growth, tumor metastasis, rheumatoid arthritis or blood-borne tumors
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumor metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
CC invention.

XX Sequence 14 AA;
SQ

Query Match 5.7%; Score 14; DB 7; Length 14;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 SCLQRTTMTPLFC 79
DB 1 SCLQRTTMTPLFC 14
|||||

RESULT 109
ADCI17508
ID ADCI17508 standard; peptide; 14 AA.

XX ADCI17508;
AC ADCI17508;
XX
XX 18-DEC-2003 (first entry)
XX
XX Type IV collagen NC1 domain related peptide SEQ ID NO:112.
DE
DE crystallised NC1 domain hexamer of type IV collagen;
KW angiogenesis inhibitor; angiogenesis-mediated disease;
KW tumor metastasis inhibitor; tumor growth inhibitor;
KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
KW antianaemic; ophthalmological; antiarteriosclerotic; antitumor;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.

XX Synthetic.
OS Homo sapiens.
OS
XX WO2003012122-A2.
PN
XX 13-FEB-2003.
PD
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
PR
XX 29-OCT-2001; 2001US-0351289P.
PR
XX 22-MAR-2002; 2002US-036854P.
PR
XX 03-JUN-2002; 2002US-0385362P.
PR

(UNIV) UNIV KANSAS MEDICAL CENT.
(SUND/) SUNDARAMOORTHY M.
(HUDS/) HUDSON B.
Sundaramoorthy M, Hudson B;
WPI; 2003-332730/31.
New polypeptide, useful for treating an angiogenesis-mediated disease or
condition consisting of glaucoma or blood-borne tumors or for inhibiting
basal lamina membrane formation in cell or tissue development.
Disclosure; SEQ ID NO 112; 169pp; English.

The present invention describes a crystallised NC1 domain hexamer of type
IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
pharmaceutical composition comprising the polypeptide and a carrier; (3)
inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
disease or condition in a mammal; (5) inhibiting tumour metastasis or
growth; (5) inhibiting endothelial cell interaction with the
extracellular matrix in an animal tissue; (6) inhibiting basal lamina
membrane formation in cell or tissue development; (7) a crystal of an NC1
domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
antitumor activities, and can be used as an inhibitor of angiogenesis,
tumor growth, tumor metastasis, rheumatoid arthritis or blood-borne tumors
or for inhibiting basal lamina membrane formation in cell or tissue
development. The methods are useful for inhibiting angiogenesis in
tissue, inhibiting tumor metastasis or growth, inhibiting endothelial
cell interaction with the extracellular matrix in an animal tissue, and
identifying inhibitors of type IV collagen assembly. The present sequence
represents a peptide which is used in the exemplification of the present
invention.

Query Match 5.7%; Score 14; DB 7; Length 14;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 SFLFVQGNQRAHQ 59
DB 1 SFLFVQGNQRAHQ 14
|||||

RESULT 110
ADCI17402
ID ADCI17402 standard; peptide; 14 AA.

XX ADCI17402;
AC ADCI17402;
XX
XX 18-DEC-2003 (first entry)
XX
XX Type IV collagen NC1 domain related peptide SEQ ID NO:3.
DE
DE crystallised NC1 domain hexamer of type IV collagen;
KW angiogenesis inhibitor; angiogenesis-mediated disease;
KW tumor metastasis inhibitor; tumor growth inhibitor;
KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
KW antianaemic; ophthalmological; antiarteriosclerotic; antitumor;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.

OS Synthetic.
OS Homo sapiens.
PN WO2003012122-A2.
PD 13-FEB-2003.
XX 26-JUL-2002; 2002WO-US023763.
XX 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289P.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX (UNIV) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX Sundaramoorthy M, Hudson B;
XX WPI; 2003-332730/31.
DR New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX Claim 5; SEQ ID NO 3; 168pp; English.
XX The present invention describes a crystallised NCI domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (5) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NCI
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and
CC antiulcer activities, and can be used as an inhibitor of angiogenesis
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
XX invention.
SQ Sequence 14 AA;
Query Match 5.7%; Score 14; DB 7; Length 14;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 75 PFLFCNVNDVCNFA 88
Db 1 PFLFCNVNDVCNFA 14
RESULT 111
ADCL17667
ID ADCL17667 standard; peptide; 14 AA.
XX
AC ADCL17667;
XX
DT 18-DEC-2003 (first entry)
XX

DE Type IV collagen NCI domain related peptide SEQ ID NO:272.
XX crystallised NCI domain hexamer of type IV collagen;
KW angiogenesis inhibitor; angiogenesis-mediated disease;
KW tumour metastasis inhibitor; tumour growth inhibitor;
KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.
OS Synthetic.
OS Homo sapiens.
XX WO2003012122-A2.
PN 13-FEB-2003.
PD 26-JUL-2002; 2002WO-US023763.
PF 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289P.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX (UNIV) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX Sundaramoorthy M, Hudson B;
XX WPI; 2003-332730/31.
DR New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX Claim 5; SEQ ID NO 272; 168pp; English.
XX The present invention describes a crystallised NCI domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (5) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NCI
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and
CC antiulcer activities, and can be used as an inhibitor of angiogenesis
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
XX invention.
SQ Sequence 14 AA;
Query Match 5.7%; Score 14; DB 7; Length 14;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 75 PFLFCNVNDVCNFA 88
Db 1 PFLFCNVNDVCNFA 14
RESULT 111
ADCL17667
ID ADCL17667 standard; peptide; 14 AA.
XX
AC ADCL17667;
XX
DT 18-DEC-2003 (first entry)
XX

Qy 186 PFLECHGRCNY 199
Db 1 PFLECHGRCNY 14

RESULT 112
ADCL17605
ID ADC17605 standard; peptide; 15 AA.
XX
AC ADC17605;
XX
AC ADC17605;
XX
DT 18-DEC-2003 (first entry)
XX
XX
DE Type IV collagen NC1 domain related peptide SEQ ID NO:210.
XX
XX crystallised NC1 domain hexamer of type IV collagen;
KW angiogenesis inhibitor; angiogenesis-mediated disease;
KW tumour metastasis inhibitor; tumour growth inhibitor;
KW endothelial cell interaction inhibitor;
KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
KW antiangiogenic; ophthalmological; antiarteriosclerotic; antiulcer;
KW endothelial cell adhesion inhibitor;
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX WO2003012122-A2.
FN
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289F.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
PI
XX
XX WPI; 2003-332730/31.
DR
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT basal lamina membrane formation in cell or tissue development.
XX
XX Disclosure; SEQ ID NO 210; 168pp; English.

The present invention describes a crystallised NC1 domain hexamer of type IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a pharmaceutical composition comprising the polypeptide and a carrier; (3) inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated disease or condition in a mammal; (5) inhibiting tumour metastasis or growth; (6) inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue; (7) inhibiting basal lamina membrane formation in cell or tissue development; (8) a crystal of an NC1 domain hexamer of type IV collagen; (9) an inhibitor of type IV collagen assembly; and (9) an inhibitor of type IV collagen assembly. A crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic, antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and antiulcer activities, and can be used as an inhibitor of angiogenesis, tumour growth, tumour metastasis, endothelial cell adhesion, endothelial cell proliferation, and basal lamina assembly. A (I) polypeptide can be used for treating an angiogenesis-mediated disease or condition consisting of glaucoma, sickle cell anaemia, ulcerative colitis, psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours or for inhibiting basal lamina membrane formation in cell or tissue

CC development. The methods are useful for inhibiting angiogenesis in tissue, inhibiting tumour metastasis or growth, inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue, and identifying inhibitors of type IV collagen assembly. The present sequence CC represents a peptide which is used in the exemplification of the present CC invention.
XX
SQ Sequence 15 AA;
Query Match 5.7%; Score 14; DB 7; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.4e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 84 VCNFASRNDYSYWL 97
Db 2 VCNFASRNDYSYWL 15
RESULT 113
AAB03336
ID AAB03336 standard; peptide; 16 AA.
XX
AC AAB03336;
XX
DT 08-SEP-2000 (first entry)
XX
XX Human epitope P23.
DE
XX Human; rhesus blood group system; Rh; Rhd; Rhce; sickle cell disease;
KW thalassaemia; Rhc; Rhd; Rhd; Rhd; alloimmunisation prevention;
KW autoimmune Rh haemolytic disease; rhesus protein; immunosuppressive;
KW vaccine.
XX
XX Homo sapiens.
XX WO2000032632-A2.
XX
XX 08-JUN-2000.
XX
XX 01-DEC-1999; 99WO-GB004027.
XX
XX 01-DEC-1998; 98GB-00026378.
XX
XX (UYAB-) UNIV ABERDEEN.
PA (COMM-) COMMON SERVICES AGENCY SCOTTISH HEALTH S.
XX
XX Urbaniak SJ, Barker RN;
XX
XX WPI; 2000-412291/35.
XX
XX Composition for prevention of alloimmunization or immunosuppression of a response elicited by alloimmunization or an autoimmune hemolytic disease, comprises an epitope of a rhesus protein.
XX
XX Disclosure; Page 82; 92pp; English.
XX
XX Human blood contains the rhesus (Rh) blood group system, and humans can either be RhD positive or negative. This can lead to complications during transfusions or pregnancy if RhD negative individuals are exposed to RhD positive blood, leading to them becoming immunised to produce anti-D. The present invention relates to new human allo- and auto-reactive T-cell epitopes (AAY9760-Y99769 and AAB03201-B03337) from RhD, Rhc, Rhd, Rhd and Rhd proteins. These epitopes bind to T-cells to elicit an immune response, i.e. immunisation. These epitopes can be used as a vaccine for the prevention of alloimmunisation or immunosuppression of a response elicited by alloimmunisation or an autoimmune haemolytic disease. CC Examples of autoimmune haemolytic diseases are sickle cell disease and thalassaemia
XX
SQ Sequence 16 AA;
Query Match 5.7%; Score 14; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 9.9e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 KIISRCQVCMKRH 244
DB 3 KIISRCQVCMKRH 16

RESULT 114
AA95913
ID AA95913 standard; peptide; 21 AA.

XX AC AA95913;
XX DT 20-NOV-2000 (first entry)
XX DE Human Goodpasture antigen N-terminal peptide GPpeplA1a9.
XX KW Goodpasture antigen binding protein; GPBP; GPpepl; human;
XX KW autoimmune disease; apoptosis; cancer; tumour; diagnosis; therapy;
XX KW mutant; mutein.
XX OS Homo sapiens.
XX OS Synthetic.
XX FN WO2000050607-A2.
XX PD 31-AUG-2000.
XX PF 24-FEB-2000; 2000WO-IB000324.
XX PR 24-FEB-1999; 99US-0121483P.
XX PA (SAUS/) SAUS J.
XX PI Saus J;
XX DR WPI; 2000-572094/53.
XX FT Novel Goodpasture antigen binding proteins useful for diagnosing and
XX FT treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX PS Example 1; Page 21; 15pp; English.

XX CC The present sequence is that of GPpeplA1a9, comprising the N-terminal 21
XX CC amino acids of human Goodpasture antigen (GP) but carrying a Ser-9 to Ala
XX CC amino acid substitution. The peptide was used to characterise the
XX CC phosphorylation activity of human Goodpasture binding protein (GPBP, see
XX CC AA95900), a novel serine/threonine kinase that specifically binds to and
XX CC phosphorylates native GPpepl. The invention provides nucleic acids (see
XX CC AA50341-53) encoding GPBP recombinant vectors, host cells, encoded
XX CC polypeptides (see AA95900-11) and antibodies. It also provides methods
XX CC for detecting the presence of an autoimmune condition or apoptosis by
XX CC detecting an increase in GPBP expression, and methods for treating an
XX CC autoimmune disorder, apoptosis or a tumour by modifying GPBP expression
XX CC or activity, especially using a GP-derived peptide

XX SQ Sequence 21 AA;

Query Match 5.7%; Score 14; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTTTGRGVFT 23
DB 8 GSPATWTTTGRGVFT 21

RESULT 115
ABG79203
ID ABG79203 standard; peptide; 21 AA.
XX AC ABG79203;
XX AC

DT 15-NOV-2002 (first entry)
XX DE Human Goodpasture protein peptide, GPpeplA1a9 mutant.
XX KW Goodpasture antigen binding protein; Goodpasture syndrome; antigen;
XX KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX KW pemphigoid; lichen planus; human.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200261430-A2.
XX PD 08-AUG-2002.
XX PF 31-JAN-2002; 2002WO-EP001010.
XX PR 31-JAN-2001; 2001US-0265249P.
XX PA (SAUS/) SAUS J.
XX PI Saus J;
XX DR WPI; 2002-619280/66.
XX PT Identifying candidate compounds for treating autoimmune conditions, e.g.
XX PT Goodpasture syndrome or lupus, comprises identifying compounds that
XX PT reduce phosphorylation of, or formation of conformational isomers of,
XX PT target proteins.

XX PS Example 1; Page 26; 217pp; English.

XX CC The invention relates to identifying candidate compounds to treat an
XX CC autoimmune condition by identifying compounds that reduce phosphorylation
XX CC of a first target protein (I) which is selected from Goodpasture antigen
XX CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)
XX CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
XX CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
XX CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
XX CC conformational isomers of the second target protein (II) (selected from
XX CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic
XX CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3
XX CC NCI domain conformational isomer, which has an amino acid sequence
XX CC identical to the wild type alpha3 type IV collagen NCI domain, is
XX CC stabilised by disulphide bonds, and has a molecular weight in a non-
XX CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
XX CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on
XX CC chromosome 5q13. The method is useful for treating autoimmune conditions,
XX CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous
XX CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
XX CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)
XX CC domain (also known as the GP antigen) peptide antigen

XX SQ Sequence 21 AA;

Query Match 5.7%; Score 14; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTTTGRGVFT 23
DB 8 GSPATWTTTGRGVFT 21

RESULT 116
ADA20237
ID ADA20237 standard; peptide; 25 AA.
XX AC ADA20237;
XX AC

XX 20-NOV-2003 (first entry)
XX T7 mutant peptide related to human type IV collagen and angiogenesis.
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytosatic; gene therapy; T7 mutant peptide; mutant; muten;
XX type IV collagen alpha 3 chain; tumstatin; human.
XX Synthetic.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Misc-difference 5 /note= "Wild-type Leu substituted by Met"
FT Misc-difference 9 /note= "Wild-type Val substituted by Ile"
FT Misc-difference 11 /note= "Wild-type Asp substituted by Asn"
XX WO2003059257-A2.
XX 24-JUL-2003.
XX 20-DEC-2002/ 2002WO-US040938.
XX 21-DEC-2001/ 2001US-00032221.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX WPI; 2003-587256/55.
XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX Claim 61; Page 45; 240pp; English.
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX invention is the amino acid sequence of the mutated T7 peptide of the
XX ADA20236) and this was derived from the amino acid sequence of tumstatin,
XX which in turn was derived from the amino acid sequence of human type IV
XX collagen alpha 3 chain.
XX Sequence 25 AA;
SQ

Query Match 5.7%; Score 14; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWL 97
DB 12 VCNFASRNDYSYWL 25

RESULT 117

AA44173
ID AAY44173 standard; peptide; 12 AA.

XX
AC AAY44173;

XX 01-FEB-2000 (first entry)

XX Bovine type IV collagen alpha3 chain protein epitope motif.

XX Recombinant; bovine; alpha3 chain; type IV collagen; detection;

XX Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.

XX Bos taurus

XX US973120-A.

XX 26-OCT-1999.

XX 07-MAR-1995; 95US-00399889.

XX 30-NOV-1990; 90US-00621091.

XX (UYVA) UNIV YALE.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Reenders ST, Morrison KE;

XX WPI; 1999-610317/52.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for
XX diagnosis and treatment of Goodpasture syndrome.

XX Claim 3; Col 36; 27pp; English.

XX This sequence represents an epitope from the bovine alpha3 chain of type
XX IV collagen polypeptide (AAY44173). The collagen polypeptide chain is
XX useful for detecting Goodpasture antibodies in blood or tissue from a
XX human patient and for treating Goodpasture syndrome especially by
XX neutralising the antibodies in the blood. The polypeptides also have a
XX nephrotrophic activity

XX Sequence 12 AA;

Query Match 4.9%; Score 12; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISRCQVCWKQKH 244

DB 1 ISRCQVCWKQKH 12

RESULT 118

AA56785

ID AAY56785 standard; peptide; 12 AA.

XX AAY56785;

XX 27-MAR-2000 (first entry)

XX Human alpha3 type IV collagen C-terminal domain fragment.

XX Goodpasture syndrome; type IV collagen; alpha3 chain; human.

XX Homo sapiens.

XX US6007980-A.

XX 28-DEC-1999.

XX

```

PF 07-OCT-1999; 98US-00167364.
XX
XX 30-NOV-1990; 90US-00621091.
PR 07-MAR-1995; 95US-00399889.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA (UYA ) UNIV YALE.
XX
XX Hudson BG, Readers ST, Morrison KE;
XX
XX WPI; 2000-096371/08.
XX
XX Diagnosing and treating Goodpasture syndrome using a peptide derived from
PT type IV collagen.
XX
XX Claim 1; Col 35; 26pp; English.
XX
XX The invention provides a method of detecting Goodpasture antibodies in
CC the fluid of a patient by contacting it with a peptide comprising at most
CC 218 amino acids of the human alpha3 chain type IV collagen that contains
CC the fragment shown in AAY56795. The methods are useful for the diagnosis
CC and treatment of Goodpasture syndrome. The present sequence represents a
CC fragment from the carboxy terminal noncollagenous domain of the human
CC alpha3 chain of type IV collagen
XX
XX Sequence 12 AA;
SQ
Query Match 4.9%; Score 12; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKXKH 244
DB 1 ISRCQVCWKXKH 12
RESULT 119
AAE09485
ID AAE09485 standard; peptide; 12 AA.
XX
XX AAE09485;
XX
XX 19-NOV-2001 (first entry)
XX
XX Human alpha-3 chain of type IV collagen peptide.
XX
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome.
XX
XX Homo sapiens.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
PR 07-MAR-1995; 95US-00399889.
PR 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX Homo sapiens.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
PR 07-MAR-1995; 95US-00399889.
PR 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX WPI; 2001-540401/60.
XX
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
PT Goodpasture antibodies from bodily fluid/tissue from patient or for
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with
PT the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
XX
XX The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is human alpha chain peptide used for constructing human
CC alpha3 chain (IV) NC1 fusion protein
XX
XX Sequence 12 AA;
SQ
Query Match 4.9%; Score 12; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKXKH 244
DB 1 ISRCQVCWKXKH 12
RESULT 120
AAE09493
ID AAE09493 standard; peptide; 12 AA.
XX
XX AAE09493;
XX
XX 19-NOV-2001 (first entry)
XX
XX Human C3 alpha-3 peptide to construct alpha3(IV)NC1 protein.
XX
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome; C3 alpha-3 peptide.
XX
XX Homo sapiens.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
PR 07-MAR-1995; 95US-00399889.
PR 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX WPI; 2001-540401/60.
XX
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
PT Goodpasture antibodies from bodily fluid/tissue from patient or for
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with
PT the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
XX
XX The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is human alpha chain peptide used for constructing human
CC alpha3 chain (IV) NC1 fusion protein
XX
XX Sequence 12 AA;
SQ
Query Match 4.9%; Score 12; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKXKH 244
DB 1 ISRCQVCWKXKH 12

```

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 209 SLNPERMFRKPI 220
Db 1.SLNPERMFRKPI 12
|||||

RESULT 121
AAB97334
ID AAB97334 standard; peptide; 12 AA.
XX AC AAB97334;
XX DT 13-AUG-2001 (first entry)
XX DE Collagen IV alpha 3 domain epitope peptide #2.
XX KW B cell; toxin; antigen specific; antibody mediated disease; virucide;
KW immunosuppressive; antiinflammatory; antiallergic; antidiabetic;
KW thyromimetic; antithyroid; vasotropic; cardiant; antiulcer;
KW neuroprotective; antirheumatic; antiarthritic; dermatological;
KW ophthalmological; nephrotropic; allergy; autoimmune disorder;
KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;
KW cardiovascular disease; immunohaematologic disorder; neurologic disease;
KW gastrointestinal disease; collagen vascular disease; renal diseases;
KW pulmonary disease; infertility disorder; collagen IV;
KW Goodpasture syndrome.
XX OS Unidentified.
XX PN WO200132853-A1.
XX PD 10-MAY-2001.
XX PF 12-OCT-2000; 2000WO-US028157.
XX PR 29-OCT-1999; 99US-0162464P.
XX PA (BIOM-) INST APPLIED BIOMEDICINE.
XX PI Chaplin JW;
XX DR WPI; 2001-316435/33.
XX PS Disclosure; Page 35; 46pp; English.

CC This invention relates to a B cell clonal toxin. The toxin is made from two moieties, the first causes the toxin to be internalised by a B cell, and the second is a biologically acceptable toxin. The invention includes a method for inactivating/killing an antigen specific B cell. A target B cell is contacted with an effective amount of a B cell clonal toxin. The method is useful for selective immunosuppression in conditions characterised by the presence of an unwanted or deleterious immune response, e.g. in the treatment of antigen specific antibody mediated disease conditions. Use of the B cell clonal toxin can result in immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic; thyromimetic; antithyroid; vasotropic; cardiant; antiulcer; neuroprotective; antirheumatic; antiarthritic; dermatological; ophthalmological; and nephrotropic activity. The toxin is particularly useful for treating a host suffering from an antigen specific antibody mediated disease condition, where the antigen specific antibody is produced by an antigen-reactive B cell population present in a host. The toxin is useful for treating allergies, viral disease conditions, and autoimmune disorders. Also treated are skin diseases; autoimmune endocrinopathies; vasculitic syndromes; cardiovascular disease; immunohaematologic disorders; gastrointestinal diseases; neurologic diseases; collagen vascular diseases; renal diseases; pulmonary diseases; and infertility disorders. The present sequence represents a collagen IV

CC alpha 3 domain epitope peptide. An antibody response to this antigen is implicated in Goodpasture syndrome, a disorder which may be treated using the toxin of the invention

CC CC

XX SQ Sequence 12 AA;
Query Match 4.9%; Score 12; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 CNVNDVCNFASR 90
Db 1 CNVNDVCNFASR 12
|||||

RESULT 122
AAB97335
ID AAB97335 standard; peptide; 12 AA.
XX AC AAB97335;
XX DT 13-AUG-2001 (first entry)
XX DE Collagen IV alpha 3 domain epitope peptide #3.
XX KW B cell; toxin; antigen specific; antibody mediated disease; virucide;
KW immunosuppressive; antiinflammatory; antiallergic; antidiabetic;
KW thyromimetic; antithyroid; vasotropic; cardiant; antiulcer;
KW neuroprotective; antirheumatic; antiarthritic; dermatological;
KW ophthalmological; nephrotropic; allergy; autoimmune disorder;
KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;
KW cardiovascular disease; immunohaematologic disorder; neurologic disease;
KW gastrointestinal disease; collagen vascular disease; renal diseases;
KW pulmonary disease; infertility disorder; collagen IV;
KW Goodpasture syndrome.
XX OS Unidentified.
XX PN WO200132853-A1.
XX PD 10-MAY-2001.
XX PF 12-OCT-2000; 2000WO-US028157.
XX PR 29-OCT-1999; 99US-0162464P.
XX PA (BIOM-) INST APPLIED BIOMEDICINE.
XX PI Chaplin JW;
XX DR WPI; 2001-316435/33.
XX PS Disclosure; Page 35; 46pp; English.

CC This invention relates to a B cell clonal toxin. The toxin is made from two moieties, the first causes the toxin to be internalised by a B cell, and the second is a biologically acceptable toxin. The invention includes a method for inactivating/killing an antigen specific B cell. A target B cell is contacted with an effective amount of a B cell clonal toxin. The method is useful for selective immunosuppression in conditions characterised by the presence of an unwanted or deleterious immune response, e.g. in the treatment of antigen specific antibody mediated disease conditions. Use of the B cell clonal toxin can result in immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic; thyromimetic; antithyroid; vasotropic; cardiant; antiulcer; neuroprotective; antirheumatic; antiarthritic; dermatological; ophthalmological; and nephrotropic activity. The toxin is particularly useful for treating a host suffering from an antigen specific antibody

CC mediated disease condition, where the antigen specific antibody is
 CC produced by an antigen-reactive B cell population present in a host. The
 CC toxin is useful for treating allergies, viral diseases conditions, and
 CC autoimmune disorders. Also treated are skin diseases; autoimmune
 CC endocrinopathies; vasculitic syndromes; cardiovascular disease;
 CC immunohaematologic disorders; gastrointestinal diseases; neurologic
 CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;
 CC and infertility disorders. The present sequence represents a collagen IV
 CC alpha 3 domain epitope peptide. An antibody response to this antigen is
 CC implicated in Goodpasture syndrome, a disorder which may be treated using
 CC the toxin of the invention
 CC
 CC Sequence 12 AA;

Query Match 4.9%; Score 12; DB 4; Length 12;
 Best Local Similarity 100.0%; Pred. No. 8.6e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 PSFIMFTSAGSE 166
 |||||
 DB 1 PSFIMFTSAGSE 12

RESULT 123
 ID AAB97333 standard; peptide; 12 AA.
 AC AAB97333;
 XX
 DT 13-AUG-2001 (first entry)
 DE Collagen IV alpha 3 domain epitope peptide #1.
 XX
 KW B cell; toxin; antigen specific; antibody mediated disease; virucide;
 KW immunosuppressive; antiinflammatory; anti-allergic; antidiabetic;
 KW thyromimetic; antithyroid; vasotropic; cardiac; antiulcer;
 KW neuroprotective; antirheumatic; antiarthritic; dermatological;
 KW ophthalmological; nephrotropic; allergy; autoimmune disorder;
 KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;
 KW cardiovascular disease; immunohaematologic disorder; neurologic disease;
 KW gastrointestinal disease; collagen vascular disease; renal diseases;
 KW pulmonary disease; infertility disorder; collagen IV;
 KW Goodpasture syndrome.

XX Unidentified.
 XX WO200132853-A1.
 XX 10-MAY-2001.
 XX
 PF 12-OCT-2000; 2000WO-US028157.
 XX
 PR 29-OCT-1999; 99US-0162464P.
 XX
 XX (BIOW-) INST APPLIED BIOMEDICINE.
 PA
 PI Chaplin JW;
 XX
 XX WPI; 2001-316435/33.
 XX
 XX B cell clonal toxin useful for treating autoimmune disorders such as
 PT Grave's disease, myocardial infarction, Crohn's disease, multiple
 PT sclerosis, comprises a group that causes toxin to be internalized by B
 PT cell.
 XX

PS Disclosure; Page 35; 46pp; English.
 XX
 XX This invention relates to a B cell clonal toxin. The toxin is made from
 CC two moieties, the first causes the toxin to be internalised by a B cell,
 CC and the second is a biologically acceptable toxin. The invention includes
 CC a method for inactivating/killing an antigen specific B cell. A target B
 CC cell is contacted with an effective amount of a B cell clonal toxin. The
 CC method is useful for selective immunosuppression in conditions

CC characterised by the presence of an unwanted or deleterious immune
 CC response, e.g. in the treatment of antigen specific antibody mediated
 CC disease conditions. Use of the B cell clonal toxin can result in
 CC immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic
 CC ; thyromimetic; antithyroid; vasotropic; cardiac; antiulcer;
 CC neuroprotective; antirheumatic; antiarthritic; dermatological;
 CC ophthalmological; and nephrotropic activity. The toxin is particularly
 CC useful for treating a host suffering from an antigen specific antibody
 CC mediated disease condition, where the antigen specific antibody is
 CC produced by an antigen-reactive B cell population present in a host. The
 CC toxin is useful for treating allergies, viral disease conditions, and
 CC autoimmune disorders. Also treated are skin diseases; autoimmune
 CC immunohaematologic disorders; gastrointestinal diseases; neurologic
 CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;
 CC and infertility disorders. The present sequence represents a collagen IV
 CC alpha 3 domain epitope peptide. An antibody response to this antigen is
 CC implicated in Goodpasture syndrome, a disorder which may be treated using
 CC the toxin of the invention
 CC
 CC Sequence 12 AA;

Query Match 4.9%; Score 12; DB 4; Length 12;
 Best Local Similarity 100.0%; Pred. No. 8.6e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 SGSPATWTTRGF 20
 |||||
 DB 1 SGSPATWTTRGF 12

RESULT 124
 ID ABP58056 standard; peptide; 12 AA.
 AC ABP58056;
 XX
 DT 03-MAR-2003 (first entry)
 DE Peptide used in endothelial cell tube formation inhibition assay.
 XX
 KW Angiogenesis; inhibitor; collagen; cytostatic; antiinflammatory;
 KW immunosuppressive; antiarthritic; antiarteriosclerotic; osteopathic;
 KW antirheumatic; ophthalmological.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 XX Modified-site 1 /note= "N-terminal acetylation"
 FT
 FT
 XX WO200266512-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 15-FEB-2002; 2002WO-US005211.
 XX
 XX 16-FEB-2001; 2001US-0269537P.
 PR 14-SEP-2001; 2001US-0322047P.
 XX

XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA
 XX Scialdome MA, Mousa SA, Shuey SW;
 PI
 XX WPI; 2003-111767/10.
 DR
 XX

XX New angiogenesis-inhibitory tripeptide useful for inhibiting endothelial
 PT cell tube formation in angiogenesis-dependent diseases such as cancer,
 PT ocular neovascularization and inflammatory diseases.
 XX
 PS Example 1; Page 27; 48pp; English.

XX The present sequence is that of an acetylated peptide including a Ser-Asn
 CC

CC -Ser tripeptide sequence. The peptide was prepared using a standard solid
 CC phase synthesis protocol using Fmoc chemistry. Inhibition of human
 CC fibroblast growth factor basic (FGF2)-stimulated human umbilical vein
 CC endothelial cells (HUVEC) by the peptide was determined in an example
 CC from the invention. Inhibition of EC tube formation by the peptide (0.015
 CC umol) was 55 +/- 2% (area) and 49 +/- 1% (length). Corresponding values
 CC for the tripeptide acetyl-Ser-Asn-Ser-carboxamide were 97 +/- 8% and 84
 CC +/- 9%, respectively, showing the tripeptide to be a more potent
 CC inhibitor of angiogenesis than the larger peptide. The invention provides
 CC methods and compositions for inhibiting endothelial cell tube formation,
 CC the initial step of tumor angiogenesis. Tripeptides, preferably SNS (Ser
 CC -Asn-Ser) or SOS (Ser-Gln-Ser), are used to inhibit angiogenesis-mediated
 CC processes such as ocular neovascular diseases, choroidal neovascular
 CC diseases, retina neovascular diseases, neovascularization of the angle,
 CC Bartorellitis, chronic inflammation, osteoarthritis, rheumatoid
 CC arthritis, atherosclerosis phlegmipoid, trachoma, or Osler-Webber-Rendu
 CC disease (all claimed). They are also useful for treating cancer,
 CC inflammatory disorders and autoimmune diseases

XX Sequence 12 AA;

Query Match 4.9%; Score 12; DB 6; Length 12;
 Best Local Similarity 100.0%; Pred. No. 8.6e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 CNYYSNSYSFWL 207
 DB 1 CNYYSNSYSFWL 12
 |||||

RESULT 125

ADCL17409
 ID ADCL17409 standard; peptide; 12 AA.

AC ADCL17409;

XX 18-DEC-2003 (first entry)

DE Type IV collagen NCI domain related peptide SEQ ID NO:10.

XX crystallised NCI domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

XX Synthetic.
 OS Homo sapiens.

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

PR 27-JUL-2001; 2001US-0308523P.

PR 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.

PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

PI Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX

PT New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.

PS Claim 10; SEQ ID NO 10; 168pp; English.

XX The present invention describes a crystallised NCI domain hexamer of type
 CC IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (5) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NCI
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NCI domain hexamer of type IV collagen (1) has cytostatic,
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.

XX Sequence 12 AA;

Query Match 4.9%; Score 12; DB 7; Length 12;
 Best Local Similarity 100.0%; Pred. No. 8.6e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 PFLECHGRGTCN 197
 DB 1 PFLECHGRGTCN 12
 |||||

RESULT 126

ADCL17634

ID ADCL17634 standard; peptide; 11 AA.

XX ADCL17634;

XX 18-DEC-2003 (first entry)

DE Type IV collagen NCI domain related peptide SEQ ID NO:239.

XX crystallised NCI domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.

XX Synthetic.

OS Homo sapiens.

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX

```

PR 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289P.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Claim 55; SEQ ID NO 239; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (1) has cytostatic,
XX antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (1) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 11 AA;
XX
XX Query Match 4.5%; Score 11; DB 7; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.00085;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 210 LNPFRMRKPI 220
XX |||||
XX 1 LNPFRMRKPI 11
XX
XX RESULT 127
XX ADC17571
XX ID ADC17571 standard; peptide; 11 AA.
XX
XX AC ADC17571;
XX
XX AC ADC17571;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX Type IV collagen NCI domain related peptide SEQ ID NO:176.
XX
XX crystallised NCI domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
XX tumour metastasis inhibitor; tumour growth inhibitor;
XX endothelial cell interaction inhibitor;
XX basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
XX antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX endothelial cell adhesion inhibitor;

```

```

KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.
XX
XX Synthetic.
XX
XX Homo sapiens.
XX
XX WO2003012122-A2.
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
XX 29-OCT-2001; 2001US-0351289P.
XX 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Claim 43; SEQ ID NO 176; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (1) has cytostatic,
XX antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (1) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 11 AA;
XX
XX Query Match 4.5%; Score 11; DB 7; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.00085;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 189 ECHGRGTCNY 199
XX |||||
XX 1 ECHGRGTCNY 11
XX
XX RESULT 128
XX ADC17545
XX ID ADC17545 standard; peptide; 11 AA.
XX
XX

```


AC ADCL17545;
 XX 18-DEC-2003 (first entry)
 DT Type IV collagen NC1 domain related peptide SEQ ID NO:150.
 XX
 XX
 XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antiprosiatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 XX Synthetic.
 OS Homo sapiens.
 OS
 XX WO2003012122-A2.
 XX
 XX 13-FEB-2003.
 PD
 XX 26-JUL-2002; 2002WO-US023763.
 PF
 XX 27-JUL-2001; 2001US-0308523P.
 XX 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX Sundaramoorthy M, Hudson B;
 FI WPI; 2003-332730/31.
 DR
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 XX Claim 36; SEQ ID NO 150; 168pp; English.
 PS
 XX The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antiprosiatic, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 XX Sequence 11 AA;
 SQ

Query Match 4.5%; Score 11; DB 7; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00085;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 102 LMPNMAPITG 112
 Db 1 LMPNMAPITG 11
 RESULT 129
 ADCL17666
 ID ADCL17666 standard; peptide; 14 AA.
 XX
 AC ADCL17666;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 XX Type IV collagen NC1 domain related peptide SEQ ID NO:271.
 DE
 XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antiprosiatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 XX Synthetic.
 OS Homo sapiens.
 OS
 XX WO2003012122-A2.
 XX
 XX 13-FEB-2003.
 PD
 XX 26-JUL-2002; 2002WO-US023763.
 PF
 XX 27-JUL-2001; 2001US-0308523P.
 XX 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX Sundaramoorthy M, Hudson B;
 FI WPI; 2003-332730/31.
 DR
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
 PT basal lamina membrane formation in cell or tissue development.
 XX
 XX Claim 57; SEQ ID NO 271; 168pp; English.
 PS
 XX The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (6) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antiprosiatic, antianaemic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be

CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 XX Sequence 14 AA;
 SQ
 Query Match 4.5%; Score 11; DB 7; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 189 ECHGRGTCNY 199
 Db 4 ECHGRGTCNY 14
 RESULT 130
 ADA20240
 ID ADA20240 standard; peptide; 19 AA.
 AC ADA20240;
 XX
 XX 20-NOV-2003 (first entry)
 DT
 XX TP3 peptide related to human type IV collagen alpha and angiogenesis.
 DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 XX metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytosolic; gene therapy; TP3 peptide; tumstatin; human;
 KW type IV collagen alpha 3 chain; mutant; mutein.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1 /note= "Wild-type Phe substituted by Lys"
 FT Misc-difference 8 /note= "Wild-type Asp substituted by Cys"
 FT
 XX WO2003059257-A2.
 PN
 XX 24-JUL-2003.
 PD
 XX 20-DEC-2002; 2002WO-US040938.
 XX
 XX 21-DEC-2001; 2001US-00032221.
 XX
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA
 XX Kalluri R;
 PI
 XX WPI; 2003-587256/55.
 DR
 XX New peptide, useful for preparing a composition for inhibiting tumor
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 PT
 XX Claim 64; Page 45; 240pp; English.
 PS
 XX This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence of the TP3 peptide of the invention,
 CC derived from the amino acid sequence of tumstatin, which in turn was
 CC derived from the amino acid sequence of human type IV collagen alpha 3
 CC chain.
 XX
 XX Sequence 19 AA;
 SQ
 Query Match 4.5%; Score 11; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 84 VCNFASRNDYS 94
 Db 9 VCNFASRNDYS 19
 RESULT 131
 ADC17413
 ID ADC17413 standard; peptide; 22 AA.
 XX
 XX ADC17413;
 AC
 XX 18-DEC-2003 (first entry)
 DT
 XX Type IV collagen NC1 domain related peptide SEQ ID NO:14.
 DE crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytosolic; antipsoriatic;
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1..5 /note= "optionally serially deleted"
 FT Misc-difference 18..22 /note= "optionally serially deleted"
 FT
 XX WO2003012122-A2.
 PN
 XX 13-FEB-2003.
 PD
 XX 26-JUL-2002; 2002WO-US023763.
 XX
 XX 27-JUL-2001; 2001US-0308523P.
 PR 29-OCT-2001; 2001US-0351289P.
 PR 22-MAR-2002; 2002US-036854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HDS/) HUDSON B.
 XX Sundaramoorthy M, Hudson B;
 PI
 XX

```

DR  WPI; 2003-332730/31.
XX
PT  New polypeptide, useful for treating an angiogenesis-mediated disease or
PT  condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT  basal lamina membrane formation in cell or tissue development.
XX
XX
PS  Claim 11; SEQ ID NO 14; 168pp; English.
XX
XX  The present invention describes a crystallised NCI domain hexamer of type
CC  IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC  pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC  inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC  disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC  growth; (5) inhibiting endothelial cell interaction with the
CC  extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC  membrane formation in cell or tissue development; (7) a crystal of an NCI
CC  domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC  collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC  crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
CC  antiulcer activities, and can be used as an inhibitor of angiogenesis,
CC  antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
CC  tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC  cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC  used for treating an angiogenesis-mediated disease or condition
CC  consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC  psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors
CC  or for inhibiting basal lamina membrane formation in cell or tissue
CC  development. The methods are useful for inhibiting angiogenesis in
CC  tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC  cell interaction with the extracellular matrix in an animal tissue, and
CC  identifying inhibitors of type IV collagen assembly. The present sequence
CC  represents a peptide which is used in the exemplification of the present
CC  invention.
XX
SQ  Sequence 22 AA;

Query Match      4.5%; Score 11; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  189 ECHGRGTCNY 199
DB  9 ECHGRGTCNY 19

RESULT 132
ADCI17415
ID  ADCI17415 standard; peptide; 22 AA.
XX
AC  ADCI17415;
XX
DT  18-DEC-2003 (first entry)
XX
XX  Type IV collagen NCI domain related peptide SEQ ID NO:16.
XX  crystallised NCI domain hexamer of type IV collagen;
XX  angiogenesis inhibitor; angiogenesis-mediated disease;
XX  tumour metastasis inhibitor; tumour growth inhibitor;
XX  endothelial cell interaction inhibitor;
XX  basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
XX  antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX  endothelial cell adhesion inhibitor;
XX  endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
XX  ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX  blood-borne tumour.
XX
OS  Synthetic.
OS  Homo sapiens.
XX
XX  Key Location/Qualifiers
XX
FT  Misc-difference 1..5
FT  Misc-difference 18..22
  /note= "optionally serially deleted"

```

/note= "optionally serially deleted"

```

FT  WPI; 2003-332730/31.
XX
XX  New polypeptide, useful for treating an angiogenesis-mediated disease or
XX  condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX  basal lamina membrane formation in cell or tissue development.
XX
XX
PF  Claim 11; SEQ ID NO 14; 168pp; English.
XX
XX  The present invention describes a crystallised NCI domain hexamer of type
XX  IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX  pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX  inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX  disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX  growth; (5) inhibiting endothelial cell interaction with the
XX  extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX  membrane formation in cell or tissue development; (7) a crystal of an NCI
XX  domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX  collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX  crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
XX  antiulcer activities, and can be used as an inhibitor of angiogenesis,
XX  antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX  tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX  cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX  used for treating an angiogenesis-mediated disease or condition
XX  consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX  psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors
XX  or for inhibiting basal lamina membrane formation in cell or tissue
XX  development. The methods are useful for inhibiting angiogenesis in
XX  tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX  cell interaction with the extracellular matrix in an animal tissue, and
XX  identifying inhibitors of type IV collagen assembly. The present sequence
XX  represents a peptide which is used in the exemplification of the present
XX  invention.
XX
SQ  Sequence 22 AA;

Query Match      4.5%; Score 11; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  189 ECHGRGTCNY 199
DB  9 ECHGRGTCNY 19

RESULT 133
ABB64070
ID  ABB64070 standard; protein; 1940 AA.
XX
XX  ABB64070;
XX
XX  26-MAR-2002 (first entry)
XX
XX  Drosophila melanogaster polypeptide SEQ ID NO 19002.
DE

```

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
PD
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
PR (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL08173.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 19002; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1940 AA;
SQ
Query Match 4.5%; Score 11; DB 4; Length 1940;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 174 SPGSCLEEFRA 184
DB 1670 SPGSCLEEFRA 1680
RESULT 134
AAG93938
ID AAG93938 standard; peptide; 10 AA.
AC AAG93938;
XX
XX 18-SEP-2001 (first entry)
DT
DE Human complementary peptide; ligand; drug discovery; drug design.
XX
XX Human; complementary peptide; ligand; drug discovery; drug design.
XX Homo sapiens.
OS
XX WO200142277-A2.
XX
PD 14-JUN-2001.
XX
XX 13-DEC-2000; 2000WO-GB004776.
XX
XX 13-DEC-1999; 99GB-00029454.
XX (PROT-) PROTEOM LTD.
PA

XX Roberts GW, Heal JR;
PI WPI; 2001-408419/43.
XX
XX A set of peptide ligands consisting of specific complementary peptides to
PT proteins encoded by genes of the human genome, useful in an assay for
PT screening and identifying of one or more novel peptides which are drug
PT candidates or pro-drugs.
XX
XX Example 4; Page 59; 646pp; English.
PS
XX The invention relates to a set of complementary peptide ligands generated
CC from the human genome. The complementary peptides interact with their
CC relevant target proteins encoded in the human genome. They can be used as
CC reagents in drug discovery and as lead ligands to facilitate drug design
CC and development. The present sequence is a complementary peptide provided
CC in the specification
XX
XX Sequence 10 AA;
SQ
Query Match 4.1%; Score 10; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0082;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 158 IMFTSAGSEGE 167
DB 1 IMFTSAGSEGE 10
RESULT 135
AAB97338
ID AAB97338 standard; peptide; 10 AA.
XX
XX AAB97338;
AC
XX 13-AUG-2001 (first entry)
DT
DE Collagen IV alpha 3 domain epitope peptide #6.
XX
XX B cell; toxin; antigen specific; antibody mediated disease; virucide;
KW immunosuppressive; antiinflammatory; antiallergic; antidiabetic;
KW thyromimetic; antithyroid; vasotropic; cardiac; antitumor;
KW neuroprotective; antirheumatic; antiarthritic; dermatological;
KW ophthalmological; nephrotropic; allergy; autoimmune disorder;
KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;
KW cardiovascular disease; immunohaematologic disorder; neurologic disease;
KW gastrointestinal disease; collagen vascular disease; renal diseases;
KW pulmonary disease; infertility disorder; collagen IV;
KW Goodpasture syndrome.
XX
XX Unidentified.
OS
XX WO200132853-A1.
XX
XX 10-MAY-2001.
PD
XX 12-OCT-2000; 2000WO-US028157.
PF
XX 29-OCT-1999; 99US-0162464P.
PR
XX (BION-) INST APPLIED BIOMEDICINE.
XX
XX Chaplin JW;
PI
XX WPI; 2001-316435/33.
DR
XX B cell clonal toxin useful for treating autoimmune disorders such as
PT Grave's disease, myocardial infarction, Crohn's disease, multiple
PT sclerosis, comprises a group that causes toxin to be internalized by B
PT cell.
XX
XX Disclosure; Page 35; 46pp; English.
PS

XX This invention relates to a B cell clonal toxin. The toxin is made from
 CC two moieties, the first causes the toxin to be internalised by a B cell,
 CC and the second is a biologically acceptable toxin. The invention includes
 CC a method for inactivating/killing an antigen specific B cell. A target B
 CC cell is contacted with an effective amount of a B cell clonal toxin. The
 CC method is useful for selective immunosuppression in conditions
 CC characterised by the presence of an unwanted or deleterious immune
 CC response, e.g. in the treatment of antigen specific antibody mediated
 CC disease conditions. Use of the B cell clonal toxin can result in
 CC immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic
 CC neuroprotective; antithyroid; vasotropic; cardiatic; antitumor;
 CC ophthalmological; and nephrotropic activity. The toxin is particularly
 CC useful for treating a host suffering from an antigen specific antibody
 CC mediated disease condition, where the antigen specific antibody is
 CC produced by an antigen-reactive B cell population present in a host. The
 CC toxin is useful for treating allergies, viral disease conditions, and
 CC autoimmune disorders. Also treated are skin diseases; autoimmune
 CC endocrinopathies; vasculitic syndromes; cardiovascular disease;
 CC immunohaematologic disorders; gastrointestinal diseases; neurologic
 CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;
 CC and infertility disorders. The present sequence represents a collagen IV
 CC alpha 3 domain epitope peptide. An antibody response to this antigen is
 CC implicated in Goodpasture syndrome, a disorder which may be treated using
 CC the toxin of the invention
 XX Sequence 10 AA;
 SQ
 Query Match 4.1%; Score 10; DB 4; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 45 FSLFLVQGNQ 54
 |||||
 Db 1 FSLFLVQGNQ 10
 RESULT 136
 AAB97336
 ID AAB97336 standard; peptide; 10 AA.
 AC AAB97336;
 XX
 DT 13-AUG-2001 (first entry)
 XX
 DE Collagen IV alpha 3 domain epitope peptide #4.
 XX
 KW B cell; toxin; antigen specific; antibody mediated disease; virucide;
 KW immunosuppressive; antiinflammatory; antiallergic; antidiabetic;
 KW neuroprotective; antithyroid; vasotropic; cardiatic; antitumor;
 KW ophthalmological; and nephrotropic activity. The toxin is particularly
 KW useful for treating a host suffering from an antigen specific antibody
 KW mediated disease condition, where the antigen specific antibody is
 KW produced by an antigen-reactive B cell population present in a host. The
 KW toxin is useful for treating allergies, viral disease conditions, and
 KW autoimmune disorders. Also treated are skin diseases; autoimmune
 KW endocrinopathies; vasculitic syndromes; cardiovascular disease;
 KW immunohaematologic disorders; gastrointestinal diseases; neurologic
 KW diseases; collagen vascular diseases; renal diseases; pulmonary diseases;
 KW and infertility disorders. The present sequence represents a collagen IV
 KW alpha 3 domain epitope peptide. An antibody response to this antigen is
 KW implicated in Goodpasture syndrome, a disorder which may be treated using
 KW the toxin of the invention
 XX Unidentified.
 OS
 XX WO200132853-A1.
 PN
 PD 10-MAY-2001.
 XX
 PF 12-OCT-2000; 2000WO-US028157.
 XX
 PR 29-OCT-1999; 99US-0162464P.
 XX
 PA (BIOM-) INST APPLIED BIOMEDICINE.
 XX
 PI Chaplin JW;
 XX

DR WPI; 2001-316435/33.
 XX B cell clonal toxin useful for treating autoimmune disorders such as
 PT Grave's disease, myocardial infarction, Crohn's disease, multiple
 PT sclerosis, comprises a group that causes toxin to be internalized by B
 PT cell.
 XX Disclosure; Page 35; 46pp; English.
 XX This invention relates to a B cell clonal toxin. The toxin is made from
 CC two moieties, the first causes the toxin to be internalised by a B cell,
 CC and the second is a biologically acceptable toxin. The invention includes
 CC a method for inactivating/killing an antigen specific B cell. A target B
 CC cell is contacted with an effective amount of a B cell clonal toxin. The
 CC method is useful for selective immunosuppression in conditions
 CC characterised by the presence of an unwanted or deleterious immune
 CC response, e.g. in the treatment of antigen specific antibody mediated
 CC disease conditions. Use of the B cell clonal toxin can result in
 CC immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic
 CC neuroprotective; antithyroid; vasotropic; cardiatic; antitumor;
 CC ophthalmological; and nephrotropic activity. The toxin is particularly
 CC useful for treating a host suffering from an antigen specific antibody
 CC mediated disease condition, where the antigen specific antibody is
 CC produced by an antigen-reactive B cell population present in a host. The
 CC toxin is useful for treating allergies, viral disease conditions, and
 CC autoimmune disorders. Also treated are skin diseases; autoimmune
 CC endocrinopathies; vasculitic syndromes; cardiovascular disease;
 CC immunohaematologic disorders; gastrointestinal diseases; neurologic
 CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;
 CC and infertility disorders. The present sequence represents a collagen IV
 CC alpha 3 domain epitope peptide. An antibody response to this antigen is
 CC implicated in Goodpasture syndrome, a disorder which may be treated using
 CC the toxin of the invention
 XX Sequence 10 AA;
 SQ
 Query Match 4.1%; Score 10; DB 4; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 207 LASLNPERMF 216
 |||||
 Db 1 LASLNPERMF 10
 RESULT 137
 ADA20257
 ID ADA20257 standard; peptide; 10 AA.
 XX
 AC ADA20257;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Peptide Seq ID58 related to human type IV collagen and angiogenesis.
 XX
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
 KW metastasis; basement membrane organisation; type IV collagen network;
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
 KW cytostatic; gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003059257-A2.
 XX
 PD 24-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-US040938.
 XX
 PR 21-DEC-2001; 2001US-00032221.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX FI Kalluri R;
 XX DR WPI; 2003-587256/55.
 XX PT New peptide, useful for preparing a composition for inhibiting tumor
 XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
 XX PS Claim 42; Page 174; 240pp; English.
 XX
 CC This invention relates to novel isolated proteins and their fragments
 CC with anti-angiogenic properties. The invention also relates to the DNA
 CC sequences which encode the novel proteins. A wide variety of diseases are
 CC the result of undesirable angiogenesis. The formation of new capillaries
 CC from pre-existing vessels is essential for tumour growth and metastasis.
 CC Basement membrane organisation is dependent on the assembly of a type IV
 CC collagen network which may occur through the C-terminal globular non-
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
 CC forms are ubiquitously exhibited in human basement membranes. In the
 CC present invention, cell surface receptors (in particular integrins) which
 CC specifically bind anti-angiogenic proteins and peptides (in particular
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA
 CC sequences of the invention may be useful in gene therapy. The present
 CC sequence is the amino acid sequence peptide Seq IDS8, derived from the
 CC sequence for human type IV collagen alpha 3 chain, which is preferably
 CC including as part of a peptide of the invention.
 XX
 SQ Sequence 10 AA;
 Query Match 4.1%; Score 10; DB 6; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 88 ASRNDYSYWL 97
 Db 1 ASRNDYSYWL 10
 |||||
 RESULT 138
 ADC17678
 ID ADC17678 standard; peptide; 10 AA.
 XX AC ADC17678;
 XX DT 18-DEC-2003 (first entry)
 XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:283.
 XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW anti-angiogenic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
 KW blood-borne tumour.
 XX OS Synthetic.
 OS Homo sapiens.
 XX WO2003012122-A2.
 XX 13-FEB-2003.
 XX 26-JUL-2002; 2002WO-US023763.
 XX 27-JUL-2001; 2001US-0308523P.
 XX 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.
 PR 03-JUN-2002; 2002US-0385362P.
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (SUND/) SUNDARAMOORTHY M.
 PA (HUDS/) HUDSON B.
 XX Sundaramoorthy M, Hudson B;
 PI WPI; 2003-332730/31.
 XX
 CC New polypeptide, useful for treating an angiogenesis-mediated disease or
 CC condition consisting of glaucoma or blood-borne tumors or for inhibiting
 CC basal lamina membrane formation in cell or tissue development.
 XX
 PS Claim 57; SEQ ID NO 283; 169pp; English.
 XX
 CC The present invention describes a crystallised NC1 domain hexamer of type
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
 CC growth; (5) inhibiting endothelial cell interaction with the
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
 CC antipsoriatic, anti-angiogenic, ophthalmological, antiarteriosclerotic and
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
 CC used for treating an angiogenesis-mediated disease or condition
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
 CC or for inhibiting basal lamina membrane formation in cell or tissue
 CC development. The methods are useful for inhibiting angiogenesis in
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
 CC cell interaction with the extracellular matrix in an animal tissue, and
 CC identifying inhibitors of type IV collagen assembly. The present sequence
 CC represents a peptide which is used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 10 AA;
 Query Match 4.1%; Score 10; DB 7; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 GSCIQRTTM 74
 Db 1 GSCIQRTTM 10
 |||||
 RESULT 139
 ADC17691
 ID ADC17691 standard; peptide; 10 AA.
 XX AC ADC17691;
 XX DT 18-DEC-2003 (first entry)
 XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:298.
 XX crystallised NC1 domain hexamer of type IV collagen;
 KW angiogenesis inhibitor; angiogenesis-mediated disease;
 KW tumour metastasis inhibitor; tumour growth inhibitor;
 KW endothelial cell interaction inhibitor;
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
 KW anti-angiogenic; ophthalmological; antiarteriosclerotic; antiulcer;
 KW endothelial cell adhesion inhibitor;
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;

KW blood-borne tumour.
XX Synthetic.
OS Homo sapiens.
XX
XX WO2003012122-A2.
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289P.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
PI
XX
XX WPI; 2003-332730/31.
DR
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT basal lamina membrane formation in cell or tissue development.
XX
XX Claim 57; SEQ ID NO 298; 168pp; English.
XX
XX The present invention describes a crystallised NC1 domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (5) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NC1
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,
CC antiproliferative, antitumour, ophthalmological, antiarteriosclerotic and
CC anticancer activities, and can be used as an inhibitor of angiogenesis,
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (1) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 10 AA;
SQ
Query Match 4.1%; Score 10; DB 7; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 78 FCVNDVCNFF 87
DB 1 FCVNDVCNFF 10
RESULT 140
ADA20239
ID ADA20239 standard; peptide; 27 AA.
XX
XX ADA20239;
XX

DT 20-NOV-2003 (first entry)
XX
XX T8-3 peptide related to human type IV collagen alpha and angiogenesis.
XX
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytostatic; gene therapy; T8-3 peptide; tumstatin; human;
KW type IV collagen alpha 3 chain; mutant; mutein.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX Location/Qualifiers
FH Key
FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"
FT Misc-difference 12 /note= "Wild-type Cys substituted by Ser"
FT Misc-difference 18 /note= "Wild-type Cys substituted by Ser"
FT
XX
XX WO2003059257-A2.
PN
XX
XX 24-JUL-2003.
PD
XX
XX 20-DEC-2002; 2002WO-US040938.
PF
XX
XX 21-DEC-2001; 2001US-00032221.
PR
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA
XX
XX Kailuri R;
PI
XX
XX WPI; 2003-587256/55.
DR
XX
XX New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
PT
XX
XX Claim 63; Page 45; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is the amino acid sequence of the T8-3 peptide of the invention,
CC derived from the amino acid sequence of tumstatin, which in turn was
CC derived from the amino acid sequence of human type IV collagen alpha 3
CC chain.
XX
XX Sequence 27 AA;
SQ
Query Match 4.1%; Score 10; DB 6; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 QRFTMPPLF 78
DB 2 QRFTMPPLF 11

RESULT 141
ADA20241
ID ADA20241 standard; peptide; 27 AA.
XX
AC ADA20241;
XX
DT 20-NOV-2003 (first entry)
XX
DE P2 peptide related to human type IV collagen alpha and angiogenesis.
XX
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
KW metastasis; basement membrane organisation; type IV collagen network;
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;
KW cytoskeletal; gene therapy; P2 peptide; tumstatin; human;
KW type IV collagen alpha 3 chain; mutant; mutein.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 1 /note= "Wild-type Leu substituted by Lys"
FT
FT Misc-difference 12 /note= "Wild-type Cys substituted by Asp"
FT
FT Misc-difference 18 /note= "Wild-type Cys substituted by Asp"
FT
FT
XX WO2003059257-A2.
XX
XX 24-JUL-2003.
XX
XX 20-DEC-2002; 2002WO-US040938.
XX
XX 21-DEC-2001; 2001US-00032221.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2003-587256/55.
XX
XX New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
PT
XX
PS Claim 65; Page 45; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
CC with anti-angiogenic properties. The invention also relates to the DNA
CC sequences which encode the novel proteins. A wide variety of diseases are
CC the result of undesirable angiogenesis. The formation of new capillaries
CC from pre-existing vessels is essential for tumour growth and metastasis.
CC Basement membrane organisation is dependent on the assembly of a type IV
CC collagen network which may occur through the C-terminal globular non-
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
CC forms are ubiquitously exhibited in human basement membranes. In the
CC present invention, cell surface receptors (in particular integrins) which
CC specifically bind anti-angiogenic proteins and peptides (in particular
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
CC collagen) are disclosed. The proteins of the invention may inhibit tumour
CC growth, angiogenic activity in mammalian tissue or protein synthesis in
CC endothelial cells and thus may exhibit cytostatic activity. The DNA
CC sequences of the invention may be useful in gene therapy. The present
CC sequence is the amino acid sequence of the P2 peptide of the invention,
CC derived from the amino acid sequence of tumstatin, which in turn was
CC derived from the amino acid sequence of human type IV collagen alpha 3
CC chain.
XX
SQ Sequence 27 AA;
Query Match 4.1%; Score 10; DB 6; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 142
ABG70168
ID ABG70168 standard; protein; 143 AA.
XX
AC ABG70168;
XX
DT 21-OCT-2002 (first entry)
XX
DE Human prey protein for Shigella ipaC #32.
XX
KW Prey protein; ospB; ospD1; ipaD; ipaC; ipaH9.8; ospG; ospC1; Shigella;
KW shigellosis; bacillary dysentery; antibacterial; yeast two-hybrid system;
KW protein-protein interaction; SID; selected interacting domain; human.
XX
XX Homo sapiens.
OS
XX WO200257303-A2.
XX
XX 25-JUL-2002.
XX
XX 11-JAN-2002; 2002WO-EP000777.
XX
XX 12-JAN-2001; 2001US-0261130P.
XX
XX (HYBR-) HYBRIGENICS.
XX
XX Legrain P;
XX
XX WPI; 2002-599706/64.
XX
XX N-PSDB; ABSS51561.
XX
XX New complex of protein-protein interactions between a bait Shigella
PT flexneri polypeptide and a prey mammalian or human placenta polypeptide
PT for treating or preventing bacillary dysentery in a mammal or human.
XX
XX Claim 7; Page 113; 162pp; English.
XX
XX The invention relates to a complex of protein-protein interactions
CC between a Shigella flexneri polypeptide (e.g. ospB, ospD1, ipaD, ipaC,
CC ipaH9.8, ospG and ospC1) and a mammalian polypeptide defined in the
CC specification. The complexes are formed using the yeast two-hybrid
CC system. Also included are (1) a recombinant host cell expressing the
CC interactions between the Shigella flexneri polypeptide and a mammalian
CC polypeptide defined in the specification; (2) selecting a modulating
CC compound that inhibits or activates the protein-protein interactions; (3)
CC a modulating compound obtained from the method of (2); (4) a SID
CC (selected interacting domain) polypeptide or its fragment or variant
CC comprising the human polypeptides appearing as ABG70042-ABG70242; (5) a
CC SID polynucleotide or its fragment or variant comprising encoding the
CC above polypeptides a vector comprising (5); (6) a recombinant host cell
CC containing the vector; and (10) a protein chip comprising Shigella
CC flexneri polypeptide and a mammalian polypeptide defined in the
CC specification. A pharmaceutical composition comprising the compound,
CC polypeptide or polynucleotide is useful for treating or preventing
CC shigellosis (bacillary dysentery) in a human or mammal. The present
CC sequence represents a human prey protein isolated by the yeast two-hybrid
CC assay, forming a complex of the invention with a shigella protein
XX
SQ Sequence 143 AA;
Query Match 4.1%; Score 10; DB 5; Length 143;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRETTMPFLF 78
Db 2 QRETTMPFLF 11
QY 128 PAIAIAVHSQ 137
Db 27 PAIAIAVHSQ 36

Query Match 4.1%; Score 10; DB 7; Length 227;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
Db 111 PAIAIAVHSQ 120

RESULT 145
AAU75588
ID AAU75588 standard; protein; 241 AA.
AC AAU75588;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human type IV collagen alpha 2 chain.
XX
KW Human; type IV collagen alpha 2 chain; cytostatic; antiangiogenic;
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;
KW Tumstatin; angiogenesis; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 166 /note= "Encoded by gtc"
XX
FN WO200151523-A2.
XX
PD 19-JUL-2001.
XX
PF 08-JAN-2001; 2001WO-US000565.
XX
PR 07-JAN-2000; 2000US-00479118.
PR 04-APR-2000; 2000US-00543371.
PR 21-JUL-2000; 2000US-00625191.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
DR WPI; 2002-188037/24.
DR N-PSDB; ABK15362.
XX
PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX
PS Example 14; Fig 11B; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis, in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by

receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 2 chain

XX
SQ Sequence 241 AA;
Query Match 4.1%; Score 10; DB 5; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
Db 126 PAIAIAVHSQ 135

RESULT 146
AAY67946
ID AAY67946 standard; protein; 242 AA.
XX
AC AAY67946;
XX
DT 03-APR-2000 (first entry)
XX
DE Human type IV collagen alpha 2 chain protein sequence SEQ ID NO:6.
XX
KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
KW benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
KW ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;
KW myocardial angiogenesis; plaque neovascularisation; angiodioma;
KW atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
KW Contraception; Obesity.
XX
OS Homo sapiens.
XX
PN WO9965940-A1.
XX
PD 23-DEC-1999.
XX
PF 17-JUN-1999; 99WO-US013737.
XX
PR 17-JUN-1998; 98US-0089689P.
PR 25-MAR-1999; 99US-0126175P.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Kalluri R;
XX
DR WPI; 2000-037708/08.
DR N-PSDB; AA257162.
XX
PT Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or 3
XX chain of type IV collagen used in, e.g. treatment of benign tumors and
XX rheumatoid arthritis.
XX
PS Example 11; Fig 10B; 117pp; English.
XX
CC The present sequence represents the human type IV collagen alpha 2 chain.
CC The present invention describes an isolated protein chosen from the NC1
CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a

CC fragment, analogue, derivative or mutant, which has anti-angiogenic
 CC properties. The anti-angiogenic proteins, multimers and chimeras are
 CC useful for inhibiting angiogenic activity in mammalian tissue, especially
 CC for treating diseases chosen from angiogenesis-dependent cancers, benign
 CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular
 CC angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori
 CC ulcers, dialysis graft vascular access stenosis, contraception and
 CC obesity. The compositions can be used to inhibit a disease characterised
 CC by angiogenic activity, in conjunction with radiation therapy,
 CC chemotherapy or immunotherapy
 XX
 SQ Sequence 242 AA;

Query Match 4.1%; Score 10; DB 3; Length 242;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
 DB 126 PAIAIAVHSQ 135

RESULT 147
 AAY31992
 ID AAY31992 standard; protein; 258 AA.

XX AAY31992;

XX 05-JAN-2000 (first entry)

DE Type IV collagen NC1 domain alpha-2 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
 KW rheumatoid arthritis; retinal neovascularization;
 KW chorioidal neovascularization; macular degeneration;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW epidemic keratoconjunctivitis; vitamin A deficiency;
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;
 KW pterygium keratitis sicca; soggrens; acne rosacea; phlyctenulosis;
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;
 KW ulcer; herpes simplex infection; Herpes zoster infection;
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;
 KW Steven's Johnson disease; radial keratotomy; sickle cell anemia;
 KW sarcoid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu; AIDS;
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;
 KW pemphigoid.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..17 /note= "BM40 signal peptide"

FT Protein 18..258

FT Peptide 18..25 /note= "mature protein"

FT Protein 26..258 /note= "affinity tag"

FT Peptide 26..258

FT Protein /note= "NC1 alpha-2 monomer"

XX

PN WO949885-A2.

XX 07-OCT-1999.

XX 26-MAR-1999; 99WO-US006445.

XX 27-MAR-1998; 98US-0079783P.

PR 29-OCT-1998; 98US-0106170P.

XX (UNIV) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Sarraz MP;

PI WPI; 1999-601297/51.

DR N-PSDB; AA220090.

XX Inhibition of angiogenesis with non-collagenous alpha chain monomer
 PT useful for treating e.g. tumor growth or metastasis, neovascularisation,
 PT etc.

PS Disclosure; Fig 17b; 56pp; English.

XX This sequence represents a recombinant type IV collagen non-collagenous
 CC (NC1) domain alpha-2 polypeptide composed of a BM40 signal sequence
 CC (which is cleaved from the mature protein) to facilitate protein
 CC secretion, and a mature protein comprising an affinity tag (facilitates
 CC purification and identification of the material) and the alpha-1 chain
 CC monomer. The invention provides methods and kits for inhibiting
 CC angiogenesis, tumour growth and metastasis, and endothelial cell
 CC interaction with the extracellular matrix, each method comprising
 CC contacting the tumour or animal tissue with 1 or more isolated type IV
 CC collagen NC1 alpha chain monomer(s) selected from the group consisting of
 CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-
 CC 96). The monomers can be produced via recombinant protein expression. The
 CC polynucleotides and polypeptides are used to treat an angiogenesis-
 CC mediated disorder or condition, especially selected from solid and blood-
 CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
 CC neovascularization, chorioidal neovascularization, macular degeneration,
 CC corneal neovascularization, retinopathy of prematurity, corneal graft
 CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
 CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, soggrens,
 CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
 CC simplex infections, herpes zoster infections, protozoan infections,
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
 CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's
 CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,
 CC sickle cell anemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets
 CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
 CC complications, abnormal proliferation of fibrovascular tissue,
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)

XX SQ Sequence 258 AA;

Query Match 4.1%; Score 10; DB 2; Length 258;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137

DB 142 PAIAIAVHSQ 151

RESULT 148

AAY97554

ID AAY97554 standard; protein; 258 AA.

XX

AC AAY97554;
 XX 12-FEB-2001 (first entry)
 XX Human alpha2(IV)NC1 protein sequence.
 XX
 XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
 XX tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
 XX retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
 XX diabetetic retinopathy; rheumatoid arthritis; neovascularisation;
 XX muscular degeneration; corneal graft rejection; vitamin A deficiency;
 XX atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
 XX Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
 XX chronic inflammation; psoriasis; therapy; alpha2(IV)NC1.
 XX
 OS Homo sapiens.
 XX
 XX WO200059532-A1.
 XX
 XX 12-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US008678.
 XX
 XX 01-APR-1999; 99US-0127391P.
 XX (BIOS-) BIOSTRATUM INC.
 XX
 XX Brooks P, Hudson B;
 XX
 XX WPI; 2000-664962/64.
 XX N-PSDB; AAA90992.
 XX
 XX Use of antagonists of specific integrin receptors for inhibiting
 XX angiogenesis, tumor growth or metastases, or endothelial cell
 XX interactions with the extracellular matrix.
 XX
 XX Disclosure; Fig 17b; 78pp; English.
 XX
 XX This sequence is a human type IV collagen alpha chain monomer, designated
 XX alpha2(IV)NC1. The invention relates to a method for inhibiting
 XX angiogenesis, tumor growth or metastases, or endothelial cell
 XX interactions with the extracellular matrix, comprising contacting the
 XX cells or tissue with a polypeptide composition containing antagonists of
 XX specific integrin receptors. The methods and the antagonists are useful
 XX for inhibiting angiogenesis, tumor growth or metastases, or endothelial
 XX cell interaction with the extracellular matrix. The antagonists are also
 XX useful for treating diseases and conditions with accompanying undesired
 XX angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
 XX carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
 XX neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
 XX treating non-tumorigenic diseases and conditions with accompanying
 XX undesired angiogenesis, e.g. diabetetic retinopathy, rheumatoid arthritis,
 XX retinal neovascularisation, choroidal neovascularisation, muscular
 XX degeneration, corneal graft rejection, vitamin A deficiency, atopic
 XX keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
 XX sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
 XX complications, chronic inflammation or psoriasis
 XX
 SQ Sequence 258 AA;
 Query Match 4.1%; Score 10; DB 3; Length 258;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 128 PAIAIAVHSQ 137
 Db 142 PAIAIAVHSQ 151
 RESULT 149
 AAB58180
 ID AAB58180 standard; protein; 430 AA.
 XX

AC AAB58180;
 XX 14-MAR-2001 (first entry)
 XX Lung cancer associated polypeptide sequence SEQ ID 518.
 XX
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;
 XX cardioactive; immunomodulatory; muscular active; vulnerary;
 XX gastrointestinal; nephroretropic; antiinfective; gynecological;
 XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 XX proliferative disorder; wound healing; infectious disease.
 XX
 OS Homo sapiens.
 XX
 XX WO200055180-A2.
 XX
 XX 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US005918.
 XX
 XX 12-MAR-1999; 99US-0124270P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX
 XX Ruben SM;
 XX
 XX WPI; 2000-587514/55.
 XX N-PSDB; AAF18056.
 XX
 XX Lung cancer associated gene sequences, referred to as lung cancer
 XX antigens, useful for treatment, prevention, and diagnosis of disorders
 XX such as lung cancer.
 XX
 XX Claim 11; Page 1008-1010; 1425pp; English.
 XX
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 XX associated proteins represented in AAB58106 - AAB58548. Lung cancer
 XX associated proteins and polynucleotide sequences, their agonists, and
 XX antagonists may have neuroprotective; cytostatic; cardioactive; and
 XX immunomodulatory; muscular active general; vulnerary; gastrointestinal
 XX general; nephroretropic; antiinfective; gynecological; or antibacterial
 XX activity. The invention also includes antibodies specific for the protein
 XX or polynucleotide sequences. The lung cancer associated polynucleotide
 XX sequences may be used for detection of lung cancer, chromosome
 XX identification, as chromosome markers, and for numerous other diagnostic
 XX or research purposes. The proteins may be used to treat disorders such as
 XX neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
 XX cardiovascular, renal, and proliferative disorders. The proteins may also
 XX be used in the treatment of wounds and infectious diseases.
 XX Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
 XX used in the course of the invention for the identification and
 XX characterisation of the polynucleotide and protein sequences
 XX
 SQ Sequence 430 AA;
 Query Match 4.1%; Score 10; DB 3; Length 430;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 128 PAIAIAVHSQ 137
 Db 314 PAIAIAVHSQ 323
 RESULT 150
 ADD89023
 ID ADD89023 standard; protein; 459 AA.
 XX
 XX ADD89023;
 XX
 XX 29-JAN-2004 (first entry)
 XX

DE TAT263.
XX tumour-associated antigenic target polypeptide; Cytostatic; tumour;
KW cancer.
XX
XX
OS Homo sapiens.
XX WO2003057160-A2.
XX
XX PD 17-JUL-2003.
XX
XX PF 30-DEC-2002; 2002WO-US041798.
XX
XX 02-JAN-2002; 2002US-0345444P.
PR 25-JAN-2002; 2002US-0351885P.
PR 25-FEB-2002; 2002US-0360066P.
PR 05-MAR-2002; 2002US-0362004P.
PR 20-MAR-2002; 2002US-0366869P.
PR 21-MAR-2002; 2002US-0366284P.
PR 28-MAR-2002; 2002US-0368679P.
PR 19-AUG-2002; 2002US-0404809P.
PR 21-AUG-2002; 2002US-0405645P.
XX
XX (GETH) GENENTECH INC.
XX
XX Prantz G, Hillan KJ, Phillips H, Polakis P, Smith V, Spencer SD;
PI Williams PM, Wu TD, Zhang Z;
XX
XX WPI; 2003-569537/53.
DR N-PSDB; ADD89098.
XX
XX PT New antibodies against tumor-associated antigenic target polypeptide,
PT useful for treating or diagnosing tumors or cancers in mammals, e.g.
PT prostate cancer, lung cancer, prostate adenocarcinomas or renal cell
PT carcinomas.
XX
XX Claim 1; SEQ ID NO 27; 252pp; English.
XX
XX CC The present invention relates to antibodies against tumour-associated
CC antigenic target polypeptide. The antibody is useful for treating or
CC diagnosing tumors or cancers in mammals, e.g. prostate cancer, lung
CC cancer, breast cancer, colon cancer, ovarian cancer, prostate
CC adenocarcinomas, renal cell carcinomas, or pleural mesothelioma. The
CC present sequence represents a TAT polypeptide.
XX
XX SQ Sequence 459 AA;
Query Match 4.1%; Score 10; DB 7; Length 459;
Best Local Similarity 100.0%; Pred. NO. 0.26;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 128 PAIAIAVHSQ 137
Db 343 PAIAIAVHSQ 352
Search completed: April 5, 2004, 07:37:22
Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:38:59 ; Search time 45 Seconds
(without alignments)
1423.863 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRDGSQSPATWTRGF.....KAGELEKIIISRCQVCMKKRH 244

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1071436 seqs, 262597696 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : Published Applications RA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 244 | 100.0 | 244 | 14 | US-10-032-221B-10 |
| 2 | 191 | 78.3 | 191 | 14 | Sequence 10, Appl |
| 3 | 159 | 65.2 | 211 | 14 | Sequence 22, Appl |
| 4 | 159 | 65.2 | 211 | 14 | Sequence 46, Appl |
| 5 | 155 | 63.5 | 232 | 14 | Sequence 46, Appl |
| 6 | 132 | 54.1 | 132 | 14 | Sequence 304, Appl |
| 7 | 124 | 50.8 | 124 | 14 | Sequence 23, Appl |
| 8 | 120 | 49.2 | 120 | 14 | Sequence 20, Appl |
| 9 | 112 | 45.9 | 112 | 14 | Sequence 21, Appl |
| 10 | 88 | 36.1 | 88 | 14 | Sequence 24, Appl |
| 11 | 80 | 32.8 | 88 | 14 | Sequence 33, Appl |
| 12 | 79 | 32.4 | 79 | 14 | Sequence 34, Appl |
| 13 | 64 | 26.2 | 64 | 14 | Sequence 26, Appl |
| 14 | 61 | 25.0 | 68 | 14 | Sequence 25, Appl |
| 15 | 61 | 25.0 | 68 | 14 | Sequence 50, Appl |

| | | | | |
|------|------|----|---------------------|--------------------|
| 25.0 | 72 | 14 | US-10-270-877-48 | Sequence 48, Appl |
| 25.0 | 72 | 14 | US-10-270-877-52 | Sequence 52, Appl |
| 25.0 | 72 | 14 | US-10-270-837-48 | Sequence 48, Appl |
| 25.0 | 72 | 14 | US-10-270-837-52 | Sequence 52, Appl |
| 15.2 | 72 | 14 | US-10-270-877-61 | Sequence 61, Appl |
| 15.2 | 72 | 14 | US-10-270-837-61 | Sequence 61, Appl |
| 10.7 | 27 | 14 | US-10-032-221B-39 | Sequence 39, Appl |
| 10.7 | 27 | 14 | US-10-032-221B-37 | Sequence 37, Appl |
| 10.2 | 25 | 14 | US-10-032-221B-37 | Sequence 37, Appl |
| 9.0 | 22 | 14 | US-10-206-699-15 | Sequence 15, Appl |
| 9.0 | 22 | 14 | US-10-206-699-266 | Sequence 266, Appl |
| 8.6 | 21 | 14 | US-10-270-877-26 | Sequence 26, Appl |
| 8.6 | 21 | 14 | US-10-270-837-26 | Sequence 26, Appl |
| 8.6 | 21 | 14 | US-10-206-699-247 | Sequence 247, Appl |
| 8.2 | 20 | 14 | US-10-206-699-289 | Sequence 289, Appl |
| 8.2 | 20 | 14 | US-10-032-221B-28 | Sequence 28, Appl |
| 8.2 | 20 | 14 | US-10-032-221B-29 | Sequence 29, Appl |
| 8.2 | 20 | 14 | US-10-032-221B-30 | Sequence 30, Appl |
| 7.8 | 19 | 14 | US-10-032-221B-27 | Sequence 27, Appl |
| 7.8 | 19 | 14 | US-10-032-221B-31 | Sequence 31, Appl |
| 7.8 | 19 | 14 | US-10-032-221B-32 | Sequence 32, Appl |
| 7.4 | 18 | 14 | US-10-206-699-254 | Sequence 254, Appl |
| 7.4 | 18 | 14 | US-10-206-699-260 | Sequence 260, Appl |
| 7.4 | 18 | 14 | US-10-206-699-277 | Sequence 277, Appl |
| 7.0 | 16 | 14 | US-09-864-741-48095 | Sequence 48095, A |
| 7.0 | 16 | 14 | US-10-206-699-302 | Sequence 302, Appl |
| 7.0 | 16 | 14 | US-10-206-699-306 | Sequence 306, Appl |
| 7.0 | 16 | 14 | US-10-032-221B-2 | Sequence 2, Appl |
| 7.0 | 16 | 14 | US-09-925-297-496 | Sequence 496, Appl |
| 7.0 | 16 | 14 | US-09-925-302-507 | Sequence 507, Appl |
| 7.0 | 16 | 14 | US-10-372-683-8 | Sequence 8, Appl |
| 6.6 | 15 | 14 | US-10-206-699-74 | Sequence 74, Appl |
| 6.1 | 15 | 14 | US-10-206-699-53 | Sequence 53, Appl |
| 6.1 | 15 | 14 | US-10-206-699-94 | Sequence 94, Appl |
| 6.1 | 15 | 14 | US-10-206-699-137 | Sequence 137, Appl |
| 6.1 | 15 | 14 | US-10-206-699-191 | Sequence 191, Appl |
| 6.1 | 15 | 14 | US-10-206-699-212 | Sequence 212, Appl |
| 6.1 | 15 | 14 | US-10-206-699-231 | Sequence 231, Appl |
| 5.7 | 14 | 14 | US-10-206-699-3 | Sequence 3, Appl |
| 5.7 | 14 | 14 | US-10-206-699-28 | Sequence 28, Appl |
| 5.7 | 14 | 14 | US-10-206-699-112 | Sequence 112, Appl |
| 5.7 | 14 | 14 | US-10-206-699-165 | Sequence 165, Appl |
| 5.7 | 14 | 14 | US-10-206-699-272 | Sequence 272, Appl |
| 5.7 | 14 | 14 | US-10-206-699-210 | Sequence 210, Appl |
| 5.7 | 21 | 14 | US-10-270-877-27 | Sequence 27, Appl |
| 5.7 | 21 | 14 | US-10-270-837-27 | Sequence 27, Appl |
| 5.7 | 25 | 14 | US-10-032-221B-38 | Sequence 38, Appl |
| 4.9 | 12 | 14 | US-10-206-699-10 | Sequence 10, Appl |
| 4.5 | 11 | 14 | US-10-206-699-150 | Sequence 150, Appl |
| 4.5 | 11 | 14 | US-10-206-699-176 | Sequence 176, Appl |
| 4.5 | 11 | 14 | US-10-206-699-239 | Sequence 239, Appl |
| 4.5 | 14 | 14 | US-10-206-699-271 | Sequence 271, Appl |
| 4.5 | 19 | 14 | US-10-032-221B-41 | Sequence 41, Appl |
| 4.5 | 22 | 14 | US-10-206-699-14 | Sequence 14, Appl |
| 4.5 | 22 | 14 | US-10-206-699-16 | Sequence 16, Appl |
| 4.5 | 1744 | 15 | US-10-369-493-5832 | Sequence 5832, Ap |
| 4.1 | 10 | 10 | US-09-572-404B-132 | Sequence 132, Ap |
| 4.1 | 10 | 14 | US-10-206-699-283 | Sequence 283, Appl |
| 4.1 | 10 | 14 | US-10-206-699-298 | Sequence 298, Appl |
| 4.1 | 10 | 14 | US-10-032-221B-58 | Sequence 58, Appl |
| 4.1 | 27 | 14 | US-10-032-221B-40 | Sequence 40, Appl |
| 4.1 | 27 | 14 | US-10-032-221B-42 | Sequence 42, Appl |
| 4.1 | 143 | 14 | US-10-043-487-342 | Sequence 342, Appl |
| 4.1 | 227 | 14 | US-10-206-699-303 | Sequence 303, Appl |
| 4.1 | 227 | 14 | US-10-032-221B-6 | Sequence 6, Appl |
| 4.1 | 430 | 9 | US-09-925-302-518 | Sequence 518, Appl |
| 4.1 | 459 | 15 | US-10-331-496A-27 | Sequence 27, Appl |
| 4.1 | 459 | 15 | US-10-372-683-30 | Sequence 30, Appl |
| 4.1 | 1712 | 10 | US-09-961-403-9 | Sequence 9, Appl |
| 3.7 | 12 | 14 | US-10-206-699-9 | Sequence 9, Appl |
| 3.7 | 16 | 14 | US-10-206-699-76 | Sequence 76, Appl |
| 3.7 | 18 | 14 | US-10-206-699-259 | Sequence 259, Appl |
| 3.7 | 18 | 14 | US-10-206-699-261 | Sequence 261, Appl |
| 3.7 | 22 | 14 | US-10-206-699-265 | Sequence 265, Appl |

89 9 3.7 22 14 US-10-206-699-267 Sequence 267, App
90 8 3.3 8 14 US-10-032-221B-50 Sequence 50, Appl
91 8 3.3 8 14 US-10-032-221B-56 Sequence 56, Appl
92 8 3.3 14 14 US-10-032-221B-44 Sequence 44, Appl
93 8 3.3 14 14 US-10-270-837-44 Sequence 44, Appl
94 8 3.3 18 14 US-10-206-699-253 Sequence 253, App
95 8 3.3 19 14 US-10-270-837-43 Sequence 43, Appl
96 8 3.3 19 14 US-10-270-837-43 Sequence 43, Appl
97 8 3.3 20 14 US-10-206-699-288 Sequence 288, App
98 8 3.3 228 14 US-10-206-699-307 Sequence 307, App
99 8 3.3 937 15 US-10-369-493-3337 Sequence 3337, App
100 7 2.9 7 14 US-10-032-221B-49 Sequence 49, Appl
101 7 2.9 7 14 US-10-032-221B-55 Sequence 55, Appl
102 7 2.9 8 14 US-10-032-221B-51 Sequence 51, Appl
103 7 2.9 14 14 US-10-206-699-26 Sequence 26, Appl
104 7 2.9 15 14 US-10-206-699-92 Sequence 92, Appl
105 7 2.9 15 14 US-10-206-699-230 Sequence 230, App
106 7 2.9 15 14 US-10-206-699-232 Sequence 232, App
107 7 2.9 70 9 US-09-864-761-37448 Sequence 37448, A
108 7 2.9 70 9 US-09-864-761-47938 Sequence 47938, A
109 7 2.9 78 12 US-10-424-599-243322 Sequence 243322, App
110 7 2.9 84 12 US-10-424-599-182069 Sequence 182069, App
111 7 2.9 87 9 US-09-764-877-1521 Sequence 1521, App
112 7 2.9 87 15 US-10-242-515-1521 Sequence 1521, App
113 7 2.9 111 12 US-10-425-114-54455 Sequence 54455, A
114 7 2.9 117 11 US-09-461-580A-18 Sequence 18, Appl
115 7 2.9 142 9 US-09-864-761-38021 Sequence 38021, A
116 7 2.9 150 15 US-10-094-749-3238 Sequence 3238, App
117 7 2.9 174 12 US-10-425-114-47774 Sequence 47774, A
118 7 2.9 198 9 US-09-727-855B-7 Sequence 7, Appl
119 7 2.9 218 12 US-10-425-114-59362 Sequence 59362, A
120 7 2.9 221 12 US-10-267-682-96 Sequence 96, Appl
121 7 2.9 221 12 US-10-267-748-96 Sequence 96, Appl
122 7 2.9 231 14 US-10-206-699-305 Sequence 305, App
123 7 2.9 233 14 US-10-230-331-28 Sequence 28, Appl
124 7 2.9 247 12 US-10-282-122A-72960 Sequence 72960, A
125 7 2.9 251 12 US-10-282-122A-76032 Sequence 76032, A
126 7 2.9 251 12 US-10-296-115-808 Sequence 808, App
127 7 2.9 300 12 US-10-424-599-235900 Sequence 235900, App
128 7 2.9 304 12 US-09-964-956-71 Sequence 71, Appl
129 7 2.9 311 12 US-10-425-114-59795 Sequence 59795, A
130 7 2.9 314 12 US-10-425-114-71839 Sequence 71839, A
131 7 2.9 317 12 US-10-425-114-50147 Sequence 50147, A
132 7 2.9 328 12 US-10-424-599-201843 Sequence 201843, App
133 7 2.9 347 9 US-09-918-568-58 Sequence 58, Appl
134 7 2.9 352 9 US-09-860-351-2 Sequence 2, Appl
135 7 2.9 352 15 US-10-353-690-44 Sequence 44, Appl
136 7 2.9 372 9 US-09-973-963-4 Sequence 4, Appl
137 7 2.9 372 9 US-09-973-064-4 Sequence 4, Appl
138 7 2.9 372 9 US-09-973-077-4 Sequence 4, Appl
139 7 2.9 372 9 US-09-973-063-4 Sequence 4, Appl
140 7 2.9 372 9 US-09-973-964-4 Sequence 4, Appl
141 7 2.9 372 9 US-09-973-072-4 Sequence 4, Appl
142 7 2.9 372 9 US-09-972-038-4 Sequence 4, Appl
143 7 2.9 372 9 US-09-972-757-4 Sequence 4, Appl
144 7 2.9 372 9 US-09-973-965-4 Sequence 4, Appl
145 7 2.9 372 9 US-09-973-941-4 Sequence 4, Appl
146 7 2.9 372 10 US-09-986-992-2 Sequence 2, Appl
147 7 2.9 372 10 US-09-971-782-4 Sequence 4, Appl
148 7 2.9 372 15 US-10-094-749-1699 Sequence 1699, App
149 7 2.9 372 16 US-10-311-764-1 Sequence 1, Appl
150 7 2.9 373 9 US-09-925-300-1655 Sequence 1655, App
151 7 2.9 383 9 US-09-147-346-2 Sequence 2, Appl
152 7 2.9 386 12 US-10-425-114-59796 Sequence 59796, A
153 7 2.9 404 12 US-10-424-599-150766 Sequence 150766, A
154 7 2.9 421 14 US-10-103-313-450 Sequence 450, App
155 7 2.9 423 12 US-10-282-122A-73255 Sequence 73255, A
156 7 2.9 424 12 US-10-425-114-40411 Sequence 40411, A
157 7 2.9 469 12 US-10-425-114-64052 Sequence 64052, A
158 7 2.9 478 11 US-09-906-179A-178 Sequence 178, App
159 7 2.9 478 11 US-09-906-179A-230 Sequence 230, App
160 7 2.9 478 12 US-10-671-403-142 Sequence 142, App
161 7 2.9 478 12 US-10-671-419-142 Sequence 142, App

162 7 2.9 478 12 US-10-670-844-142 Sequence 142, App
163 7 2.9 478 12 US-10-671-134-142 Sequence 142, App
164 7 2.9 478 12 US-10-673-098-142 Sequence 142, App
165 7 2.9 488 12 US-10-424-599-198256 Sequence 198256, A
166 7 2.9 490 12 US-10-425-114-47211 Sequence 47211, A
167 7 2.9 503 14 US-10-096-840D-2 Sequence 2, Appl
168 7 2.9 503 14 US-10-065-133A-8 Sequence 8, Appl
169 7 2.9 565 14 US-10-065-133A-11 Sequence 11, Appl
170 7 2.9 565 14 US-10-065-133A-11 Sequence 11, Appl
171 7 2.9 565 16 US-10-434-811A-11 Sequence 11, Appl
172 7 2.9 585 12 US-10-425-114-53676 Sequence 53676, A
173 7 2.9 601 9 US-09-480-236-1 Sequence 1, Appl
174 7 2.9 613 9 US-09-480-236-3 Sequence 3, Appl
175 7 2.9 613 10 US-09-462-713-2 Sequence 2, Appl
176 7 2.9 618 9 US-09-817-676A-14 Sequence 14, Appl
177 7 2.9 618 9 US-09-970-516-4 Sequence 4, Appl
178 7 2.9 618 14 US-10-354-358-78 Sequence 78, Appl
179 7 2.9 622 9 US-09-147-346-4 Sequence 4, Appl
180 7 2.9 676 11 US-09-873-155-40 Sequence 40, Appl
181 7 2.9 676 14 US-10-096-840D-4 Sequence 4, Appl
182 7 2.9 693 12 US-10-425-114-56454 Sequence 56454, A
183 7 2.9 703 12 US-10-425-114-58817 Sequence 58817, A
184 7 2.9 731 14 US-10-094-113-6 Sequence 6, Appl
185 7 2.9 951 15 US-10-363-493-3330 Sequence 3330, App
186 7 2.9 1050 12 US-10-282-122A-67554 Sequence 67554, A
187 7 2.9 1759 15 US-10-369-493-7032 Sequence 7032, App
188 7 2.9 3939 14 US-10-156-761-10434 Sequence 10434, A
189 6 2.5 6 14 US-10-206-699-122 Sequence 122, App
190 6 2.5 6 14 US-10-032-221B-48 Sequence 48, Appl
191 6 2.5 6 14 US-10-032-221B-54 Sequence 54, Appl
192 6 2.5 7 14 US-10-080-505-58 Sequence 58, Appl
193 6 2.5 8 10 US-09-839-996-7 Sequence 7, Appl
194 6 2.5 8 10 US-09-839-996-8 Sequence 8, Appl
195 6 2.5 8 14 US-10-080-505-53 Sequence 53, Appl
196 6 2.5 8 14 US-10-080-505-54 Sequence 54, Appl
197 6 2.5 9 9 US-09-780-053-452 Sequence 452, App
198 6 2.5 9 13 US-10-056-151-82 Sequence 82, Appl
199 6 2.5 9 13 US-10-035-688-4 Sequence 4, Appl
200 6 2.5 9 14 US-10-032-221B-57 Sequence 57, Appl

ALIGNMENTS

RESULT 1

US-10-032-221B-10
; Sequence 10, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREC
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 244
; TYPE: PRT

```
; ORGANISM: Homo sapiens
US-10-032-221B-10
Query Match      100.0%; Score 244; DB 14; Length 244;
Best Local Similarity 100.0%; Pred. No. 1.9e-233;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPYSGFSLFVQGNQRAHQD 60

QY 61 LGTGLSCLORETTMPFLFCNVNDVCFNPNFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120
DB 61 LGTGLSCLORETTMPFLFCNVNDVCFNPNFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALLASPGSCLE 180
DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALLASPGSCLE 180

QY 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKIIISRCQVCM 240
DB 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKIIISRCQVCM 240

QY 241 KKRH 244
DB 241 KKRH 244

RESULT 2
US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US20030144881A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032.221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22
Query Match      78.3%; Score 191; DB 14; Length 191;
Best Local Similarity 100.0%; Pred. No. 5.4e-181;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 ORAHQDGLTGSCLORETTMPFLFCNVNDVCFNPNFASRNDYSYWLSTPALMPMNPAPITGR 113
DB 1 ORAHQDGLTGSCLORETTMPFLFCNVNDVCFNPNFASRNDYSYWLSTPALMPMNPAPITGR 60

QY 114 ALEPYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALLA 173
DB 61 ALEPYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALLA 120
```

```
QY 174 SPGSCLEBFPRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKII 233
DB 121 SPGSCLEBFPRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKII 180

QY 234 SRCQVCMKKRH 244
DB 181 SRCQVCMKKRH 191

RESULT 3
US-10-270-877-46
; Sequence 46, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46
Query Match      65.2%; Score 159; DB 14; Length 211;
Best Local Similarity 100.0%; Pred. No. 3.1e-149;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGFTVTRHSQTTPSCPEGTVPYSGFSLFVQGNQRAHQD 60

QY 61 LGTGLSCLORETTMPFLFCNVNDVCFNPNFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120
DB 61 LGTGLSCLORETTMPFLFCNVNDVCFNPNFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159
DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159

RESULT 4
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
```


US-10-270-837-46

Query Match 65.2%; Score 159; DB 14; Length 211;
Best Local Similarity 100.0%; Pred. No. 3.1e-149;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGFTVTRHSTTAIPSCPEGTVPYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGPATWTRGFTVTRHSTTAIPSCPEGTVPYSGFSLFVQGNQRAHQD 60

QY 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120
DB 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSEIM 159
DB 121 RCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSEIM 159

RESULT 5

US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 304
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 3 chain
US-10-206-699-304

Query Match 63.5%; Score 155; DB 14; Length 232;
Best Local Similarity 100.0%; Pred. No. 3.1e-145;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATWTRGFTVTRHSTTAIPSCPEGTVPYSGFSLFVQGNQRAHQDGLTGLSCLOQRT 72
DB 1 ATWTRGFTVTRHSTTAIPSCPEGTVPYSGFSLFVQGNQRAHQDGLTGLSCLOQRT 60

QY 73 TWPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGRALEPYISRCTVCEGPAIAI 132
DB 61 TWPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGRALEPYISRCTVCEGPAIAI 120

QY 133 AVHSQTTDIPPCPHGWISLWKGFSEIMFTSAGSEG 167
DB 121 AVHSQTTDIPPCPHGWISLWKGFSEIMFTSAGSEG 155

RESULT 6

US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match 54.1%; Score 132; DB 14; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.2e-122;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGFTVTRHSTTAIPSCPEGTVPYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGPATWTRGFTVTRHSTTAIPSCPEGTVPYSGFSLFVQGNQRAHQD 60

QY 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120
DB 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAI 132
DB 121 RCTVCEGPAIAI 132

RESULT 7

US-10-032-221B-20
; Sequence 20, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match      50.8%; Score 124; DB 14; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.8e-115;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLGKKGDSGSPATWTRGVFTRHSQTTAIPSCPEGTVPLYSGFSLVQGNQRAHGQD 60
   |||||
Db 1 GLGKKGDSGSPATWTRGVFTRHSQTTAIPSCPEGTVPLYSGFSLVQGNQRAHGQD 60
   |||||

QY 61 LGTIGLCQRFTHMPFLFCNVNDVCFNPNRNDYSYVLSLTPALPMNMAITGRALEPYIS 120
   |||||
Db 61 LGTIGLCQRFTHMPFLFCNVNDVCFNPNRNDYSYVLSLTPALPMNMAITGRALEPYIS 120
   |||||

QY 121 RCTV 124
   |||||
Db 121 RCTV 124
   |||||

RESULT 8
US-10-032-221B-21
; Sequence 21, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 334 (amino acids 125-244 of SEQ ID NO:10)
US-10-032-221B-21

Query Match      49.2%; Score 120; DB 14; Length 120;
Best Local Similarity 100.0%; Pred. No. 8.8e-111;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 CEGPAIAVHSQTTDIPPCPGHWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRA 184
   |||||
Db 1 CEGPAIAVHSQTTDIPPCPGHWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRA 60
   |||||

QY 185 SPLECHGRGTCNYNSYNSFWLASLNPERMFKPIPTVXAGELEKIISRCQVCMKKRH 244
   |||||
Db 61 SPLECHGRGTCNYNSYNSFWLASLNPERMFKPIPTVXAGELEKIISRCQVCMKKRH 120
   |||||

RESULT 9
US-10-032-221B-24
; Sequence 24, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-24

Query Match      45.9%; Score 112; DB 14; Length 112;
Best Local Similarity 100.0%; Pred. No. 7.1e-103;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 AVHSQTTDIPPCPGHWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRA 192
   |||||
Db 1 AVHSQTTDIPPCPGHWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRA 60
   |||||

QY 193 RGTCTNYNSYNSFWLASLNPERMFKPIPTVXAGELEKIISRCQVCMKKRH 244
   |||||
Db 61 RGTCTNYNSYNSFWLASLNPERMFKPIPTVXAGELEKIISRCQVCMKKRH 112
   |||||

RESULT 10
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 334 (amino acids 125-244 of SEQ ID NO:10)
US-10-032-221B-33
```


Query Match 26.2%; Score 64; DB 14; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.7e-55;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPFRMRKPIPTSTVKAGELEKIIISRCQVCM 240
Db 1 EFRASPFLECHGRGTCNYNSYSFWLASLNPFRMRKPIPTSTVKAGELEKIIISRCQVCM 60
QY 241 KKRH 244
Db 61 KKRH 64

RESULT 14
US-10-270-877-50
; Sequence 50, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 50
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-IV-V
US-10-270-877-50

Query Match 25.0%; Score 61; DB 14; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.7e-52;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Db 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 L 61
Db 61 L 61

RESULT 15
US-10-270-837-50
; Sequence 50, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 50
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-IV-V

US-10-270-837-50

Query Match 25.0%; Score 61; DB 14; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.7e-52;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Db 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 L 61
Db 61 L 61

RESULT 16
US-10-270-877-48
; Sequence 48, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 48
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII
US-10-270-877-48

Query Match 25.0%; Score 61; DB 14; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.7e-52;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
Db 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 L 61
Db 61 L 61

RESULT 17
US-10-270-877-52
; Sequence 52, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 52
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-V
US-10-270-837-52

Query Match 25.0%; Score 61; DB 14; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.7e-52; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60
DB 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60

QY 61 L 61
DB 61 L 61

RESULT 18

US-10-270-837-48
; Sequence 48, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII

US-10-270-837-48

Query Match 25.0%; Score 61; DB 14; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.7e-52; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60
DB 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60

QY 61 L 61
DB 61 L 61

RESULT 19

US-10-270-837-52
; Sequence 52, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-V
US-10-270-837-52

Query Match 25.0%; Score 61; DB 14; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.7e-52; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60
DB 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60

QY 61 L 61
DB 61 L 61

RESULT 20

US-10-270-877-61
; Sequence 61, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-10-270-877-61

Query Match 15.2%; Score 37; DB 14; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEG 37
DB 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCPEG 37

RESULT 21

US-10-270-837-61
; Sequence 61, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-270-837-61

Query Match 15.2%; Score 37; DB 14; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 266
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-266

Query Match      9.0%; Score 22; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.8e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTTMPFLFCNVNDVCNFSRND 92
Db 1 FTTMPFLFCNVNDVCNFSRND 22

RESULT 26
US-10-270-877-26
; Sequence 26, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/10/270,877
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US/10/270,877
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl
US-10-270-877-26

Query Match      8.6%; Score 21; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.7e-13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KGRGDSGSPATWTRGFVFT 23
Db 1 KGRGDSGSPATWTRGFVFT 21

RESULT 27
US-10-837-26
; Sequence 26, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
```

```
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl
US-10-270-837-26

Query Match      8.6%; Score 21; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.7e-13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KGRGDSGSPATWTRGFVFT 23
Db 1 KGRGDSGSPATWTRGFVFT 21

RESULT 28
US-10-206-699-247
; Sequence 247, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 247
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-247

Query Match      8.6%; Score 21; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.7e-13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 FWLASLNPERMFKRPISTVK 225
Db 1 FWLASLNPERMFKRPISTVK 21

RESULT 29
US-10-206-699-289
; Sequence 289, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
```

; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 289
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-693-289

Query Match 8.2%; Score 20; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRFTMPFLFCNVNDVCF 87
Db 1 LQRFTMPFLFCNVNDVCF 20

RESULT 30

US-10-032-221B-28
; Sequence 28, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence

; OTHER INFORMATION: T2 (amino acids 53-72 of SEQ ID NO:10)
US-10-032-221B-28

Query Match 8.2%; Score 20; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 NORAGQDLGTLGSCLOQRT 72
Db 1 NORAGQDLGTLGSCLOQRT 20

RESULT 31

US-10-032-221B-29
; Sequence 29, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T3 (amino acids 58-87 of SEQ ID NO:10)
US-10-032-221B-29

Query Match 8.2%; Score 20; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRFTMPFLFCNVNDVCF 87
Db 1 LQRFTMPFLFCNVNDVCF 20

RESULT 32

US-10-032-221B-30
; Sequence 30, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T4 (amino acids 83-102 of SEQ ID NO:10)
US-10-032-221B-30

Query Match 8.2%; Score 20; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 DVCNFSRNDYSYWLSTPAL 102
Db 1 DVCNFSRNDYSYWLSTPAL 20

RESULT 33

US-10-032-221B-27
; Sequence 27, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T1 (amino acids 1-19 of SEQ ID NO:10)
US-10-032-221B-27

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKRGDSGSPATWTRG 19
Db 1 GLKGKRGDSGSPATWTRG 19

RESULT 34

US-10-032-221B-31
; Sequence 31, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T5 (amino acids 98-116 of SEQ ID NO:10)
US-10-032-221B-31

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 STPALPMNMNMAPITGRALE 116
Db 1 STPALPMNMNMAPITGRALE 19

RESULT 35

US-10-032-221B-32
; Sequence 32, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T6 (amino acids 113-131 of SEQ ID NO:10)
US-10-032-221B-32

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPVIISRCTVCEGPAIA 131
Db 1 RALEPVIISRCTVCEGPAIA 19

RESULT 36

US-10-206-699-254
; Sequence 254, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523

; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 254
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-254

Query Match 7.4%; Score 18; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTTFPFLFCNVNDVCNFA 88
|||||
Db 1 FTTFPFLFCNVNDVCNFA 18

RESULT 37

US-10-206-699-260
; Sequence 260, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 260
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-260

Query Match 7.4%; Score 18; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PFLFCNVNDVCNFA 92
|||||
Db 1 PFLFCNVNDVCNFA 18

RESULT 38

US-10-206-699-277
; Sequence 277, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 277
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-277

Query Match 7.4%; Score 18; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 LNPFMRFRKPISTVKAG 227
|||||
Db 1 LNPFMRFRKPISTVKAG 18

RESULT 39

US-09-864-761-48095
; Sequence 48095, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48095
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035425.11
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EST HUMAN HIT: AW893189.1, EVALUE 2.00e-22
; OTHER INFORMATION: SWISSPROT HIT: P29400, EVALUE 2.00e-23
US-09-864-761-48095

Query Match 7.0%; Score 17; DB 9; Length 46;
Best Local Similarity 100.0%; Pred. No. 4.9e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100
Db 10 VCNFASRNDYSYWLSTP 26
|||||

RESULT 40

US-10-206-699-302
; Sequence 302, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 302
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 1 chain
US-10-206-699-302

Query Match 7.0%; Score 17; DB 14; Length 229;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100
Db 70 VCNFASRNDYSYWLSTP 86
|||||

RESULT 41

US-10-206-699-306
; Sequence 306, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT FILING DATE: 2002-07-26
US-10-206-699-306

; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 306
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: alpha 5 chain
US-10-206-699-306

Query Match 7.0%; Score 17; DB 14; Length 229;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100
Db 70 VCNFASRNDYSYWLSTP 86
|||||

RESULT 42

US-10-032-221B-2
; Sequence 2, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-2

Query Match 7.0%; Score 17; DB 14; Length 229;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100
Db 70 VCNFASRNDYSYWLSTP 86
|||||

RESULT 43

US-09-925-297-496
; Sequence 496, Application US/09925297
; Patent No. US20020081659A1

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 496
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (247)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-496

Query Match          7.0%; Score 17; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYSYWLSTP 100
Db      150 VCNFASRNDYSYWLSTP 166
|||||

RESULT 44
US-09-925-302-507
; Sequence 507, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 507
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-507

Query Match          7.0%; Score 17; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYSYWLSTP 100
Db      247 VCNFASRNDYSYWLSTP 263
|||||

RESULT 45
US-10-372-683-8
; Sequence 8, Application US/10372683
; Publication No. US20040009171A1
; GENERAL INFORMATION:
; APPLICANT: GERRITSEN, MARY E.
; APPLICANT: PEALE JR., FRANKLIN V.
```

```
; APPLICANT: WU, THOMAS D.
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
; FILE REFERENCE: P192831P1
; CURRENT APPLICATION NUMBER: US/10/372,683
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 10/271,690
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/344,534
; PRIOR FILING DATE: 2001-10-18
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 8
; LENGTH: 1669
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-10-372-683-8

Query Match          7.0%; Score 17; DB 15; Length 1669;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYSYWLSTP 100
Db      1510 VCNFASRNDYSYWLSTP 1526
|||||

RESULT 46
US-10-206-699-74
; Sequence 74, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCl Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-74

Query Match          6.8%; Score 16; DB 14; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      50 VQGNQRAHGQDLGTLG 65
Db      1 VQGNQRAHGQDLGTLG 16
|||||

RESULT 47
US-10-206-699-53
; Sequence 53, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCl Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
```

```
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-53

Query Match
Best Local Similarity 100.0%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SLNPERMFRKPIPT 223
Db 1 SLNPERMFRKPIPT 15

RESULT 48
US-10-206-699-94
; Sequence 94, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 94
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-94

Query Match
Best Local Similarity 100.0%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 TSAGSEGTGQALASP 175
Db 1 TSAGSEGTGQALASP 15

RESULT 49
US-10-206-699-137
; Sequence 137, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
```

```
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 137
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-137

Query Match
Best Local Similarity 100.0%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 PALMPMNWAPITGRA 114
Db 1 PALMPMNWAPITGRA 15

RESULT 50
US-10-206-699-191
; Sequence 191, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-191

Query Match
Best Local Similarity 100.0%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 FRKPIPTVKAGELE 230
Db 1 FRKPIPTVKAGELE 15

RESULT 51
US-10-206-699-212
; Sequence 212, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
```

; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-212

Query Match 6.1%; Score 15; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 DVCNPFASRNDYSYWL 97
|||||
DB 1 DVCNPFASRNDYSYWL 15

RESULT 52

US-10-206-699-231
; Sequence 231, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 231
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-231

Query Match 6.1%; Score 15; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 RGTCKYNSYSYFWL 207
|||||
DB 1 RGTCKYNSYSYFWL 15

RESULT 53

US-10-206-699-3
; Sequence 3, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-3

Query Match 5.7%; Score 14; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PFLFCNVNDVCNPA 88
|||||
DB 1 PFLFCNVNDVCNPA 14

RESULT 54

US-10-206-699-28
; Sequence 28, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-28

Query Match 5.7%; Score 14; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 SCLQRETTMPFLFC 79
|||||
DB 1 SCLQRETTMPFLFC 14

RESULT 55

US-10-206-699-112
; Sequence 112, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854

;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 112
;; LENGTH: 14
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-206-699-112

Query Match 5.7%; Score 14; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 SELFVQGNQRAHQ 59
DB 1 SELFVQGNQRAHQ 14

RESULT 56
US-10-206-699-165
;; Sequence 165, Application US/10206699
;; Publication No. US20030100510A1
;; GENERAL INFORMATION:
;; APPLICANT: Sundaramoorthy, M.
;; APPLICANT: Hudson, B.
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
;; FILE REFERENCE: MBHB 01-1017
;; CURRENT APPLICATION NUMBER: US/10/206,699
;; CURRENT FILING DATE: 2002-07-26
;; PRIOR APPLICATION NUMBER: US 60/308,523
;; PRIOR FILING DATE: 2001-07-27
;; PRIOR APPLICATION NUMBER: US 60/351,289
;; PRIOR FILING DATE: 2001-10-29
;; PRIOR APPLICATION NUMBER: US 60/366,854
;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 165
;; LENGTH: 14
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-206-699-165

Query Match 5.7%; Score 14; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 IMFTSAGSEGTGQA 171
DB 1 IMFTSAGSEGTGQA 14

RESULT 57
US-10-206-699-272
;; Sequence 272, Application US/10206699
;; Publication No. US20030100510A1
;; GENERAL INFORMATION:
;; APPLICANT: Sundaramoorthy, M.
;; APPLICANT: Hudson, B.
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
;; FILE REFERENCE: MBHB 01-1017
;; CURRENT APPLICATION NUMBER: US/10/206,699
;; CURRENT FILING DATE: 2002-07-26
;; PRIOR APPLICATION NUMBER: US 60/308,523
;; PRIOR FILING DATE: 2001-07-27
;; PRIOR APPLICATION NUMBER: US 60/351,289
;; PRIOR FILING DATE: 2001-10-29
;; PRIOR APPLICATION NUMBER: US 60/366,854
;; PRIOR FILING DATE: 2002-03-22

;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 272
;; LENGTH: 14
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-206-699-272

Query Match 5.7%; Score 14; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 PFLECHGRGTCNY 199
DB 1 PFLECHGRGTCNY 14

RESULT 58
US-10-206-699-210
;; Sequence 210, Application US/10206699
;; Publication No. US20030100510A1
;; GENERAL INFORMATION:
;; APPLICANT: Sundaramoorthy, M.
;; APPLICANT: Hudson, B.
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
;; FILE REFERENCE: MBHB 01-1017
;; CURRENT APPLICATION NUMBER: US/10/206,699
;; CURRENT FILING DATE: 2002-07-26
;; PRIOR APPLICATION NUMBER: US 60/308,523
;; PRIOR FILING DATE: 2001-07-27
;; PRIOR APPLICATION NUMBER: US 60/351,289
;; PRIOR FILING DATE: 2001-10-29
;; PRIOR APPLICATION NUMBER: US 60/366,854
;; PRIOR FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: US 60/385,362
;; PRIOR FILING DATE: 2002-06-03
;; NUMBER OF SEQ ID NOS: 307
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 210
;; LENGTH: 15
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-206-699-210

Query Match 5.7%; Score 14; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWL 97
DB 2 VCNFASRNDYSYWL 15

RESULT 59
US-10-270-877-27
;; Sequence 27, Application US/10270877
;; Publication No. US20030049791A1
;; GENERAL INFORMATION:
;; APPLICANT: Saus, Juan
;; TITLE OF INVENTION: Goodpasture Binding Protein
;; FILE REFERENCE: 98-723-AD1
;; CURRENT APPLICATION NUMBER: US/10/270,877
;; CURRENT FILING DATE: 2002-10-11
;; PRIOR APPLICATION NUMBER: 09/512,563
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: 60/121,483
;; PRIOR FILING DATE: 1993-02-24
;; NUMBER OF SEQ ID NOS: 63
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 27
;; LENGTH: 21

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T7-mutant (amino acids 73-97 of SEQ ID NO:10; methionine has been substituted for the leucine residue at position 77 of the full-length Tmstatin molecule, and isoleucine has been substituted for the valine at position 81, and asparagine has been substituted for the spartic acid at position 83)
US-10-032-221B-38

Query Match          5.7%; Score 14; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWL 97
      |||||
DB      12 VCNFASRNDYSYWL 25

RESULT 62
US-10-206-699-10
; Sequence 10, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-10

Query Match          4.9%; Score 12; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      186 PFLECHGRGTGN 197
      |||||
DB      1 PFLECHGRGTGN 12

RESULT 63
US-10-206-699-150
; Sequence 150, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpeplAla9
US-10-270-877-27

Query Match          5.7%; Score 14; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 GSPATWTRGFVFT 23
      |||||
DB      8 GSPATWTRGFVFT 21

RESULT 60
US-10-270-837-27
; Sequence 27, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpeplAla9
US-10-270-837-27

Query Match          5.7%; Score 14; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 GSPATWTRGFVFT 23
      |||||
DB      8 GSPATWTRGFVFT 21

RESULT 61
US-10-032-221B-38
; Sequence 38, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: PCT/US01/00565
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58

```


; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-150

Query Match 4.5%; Score 11; DB 14; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LMPNMNAPITG 112
|||||
Db 1 LMPNMNAPITG 11

RESULT 64

US-10-206-699-176
; Sequence 176, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 176
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-176

Query Match 4.5%; Score 11; DB 14; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYY 199
|||||
Db 1 ECHGRGTCNYY 11

RESULT 65

US-10-206-699-239
; Sequence 239, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 239
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-239

Query Match 4.5%; Score 11; DB 14; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 LNPFRMRKPI 220
|||||
Db 1 LNPFRMRKPI 11

RESULT 66

US-10-206-699-271
; Sequence 271, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 271
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-271

Query Match 4.5%; Score 11; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYY 199
|||||
Db 4 ECHGRGTCNYY 14

RESULT 67

US-10-032-221B-41
; Sequence 41, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118

; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: TP3 (amino acids 76-94 of SEQ ID NO:10; lysine has been substituted
; OTHER INFORMATION: ed for the phenylalanine residue at position 76 of the full-length
; OTHER INFORMATION: h Tumstatin molecule, and cysteine has been substituted for the
; OTHER INFORMATION: aspartic acid at position 83)
US-10-032-221B-41

Query Match 4.5%; Score 11; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYS 94
| | | | | | | | | |
Db 9 VCNFASRNDYS 19

RESULT 68
US-10-206-699-14
; Sequence 14, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(5)
; OTHER INFORMATION: Amino acids at positions 1-5 are optionally absent, such that if
; OTHER INFORMATION: 5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(22)
; OTHER INFORMATION: Amino acids at positions 18-22 are optionally absent, such that
; OTHER INFORMATION: 18 is absent, 19-22 are absent, if 19 is absent, 20-22 are absent

US-10-206-699-15
Query Match 4.5%; Score 11; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 ECHGRGTCNYY 199
| | | | | | | | | |
Db 9 ECHGRGTCNYY 19

RESULT 70
US-10-369-493-5832
; Sequence 5832, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5832
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5832

Query Match 4.5%; Score 11; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 ECHGRGTCNYY 199
| | | | | | | | | |
Db 9 ECHGRGTCNYY 19

RESULT 69
US-10-206-699-16
; Sequence 16, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(5)
; OTHER INFORMATION: Amino acids at positions 1-5 are optionally absent, such that if
; OTHER INFORMATION: 5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(22)
; OTHER INFORMATION: Amino acids at positions 18-22 are optionally absent, such that
; OTHER INFORMATION: 18 is absent, 19-22 are absent, if 19 is absent, 20-22 are absent

US-10-206-699-16
Query Match 4.5%; Score 11; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 ECHGRGTCNYY 199
| | | | | | | | | |
Db 9 ECHGRGTCNYY 19

RESULT 70
US-10-369-493-5832
; Sequence 5832, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5832
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5832

Query Match 4.5%; Score 11; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 ECHGRGTCNYY 199
| | | | | | | | | |
Db 9 ECHGRGTCNYY 19

RESULT 70
US-10-369-493-5832
; Sequence 5832, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5832
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5832

Query Match 4.5%; Score 11; DB 15; Length 1744;

Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLEEFRA 184
| | | | | | | | | |
DB 1675 SPGSCLEEFRA 1685

RESULT 71

US-09-572-404B-132
; Sequence 132, Application US/09572404B
; Publication No. US20030078374A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Complementary peptide ligands from the human genome
; FILE REFERENCE: Human patent
; CURRENT APPLICATION NUMBER: US/09/572,404B
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 4203
; SOFTWARE: ProtPatent version 1.0
; SEQ ID NO 132
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: sequence located in COL11A2 at 1584-1593 and may interact with Se
; OTHER INFORMATION: 131 in this patent.
US-09-572-404B-132

Query Match 4.1%; Score 10; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;

QY 158 IMFTSAGSEG 167
| | | | | | | | | |
DB 1 IMFTSAGSEG 10

RESULT 72

US-10-206-699-283
; Sequence 283, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 283
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-283

Query Match 4.1%; Score 10; DB 14; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLQRFTTM 74
| | | | | | | | | |
DB 1 GSCLQRFTTM 10

RESULT 73

US-10-206-699-298
; Sequence 298, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 298
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-298

Query Match 4.1%; Score 10; DB 14; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;

QY 78 FCNVNDVCNF 87
| | | | | | | | | |
DB 1 FCNVNDVCNF 10

RESULT 74

US-10-032-221B-58
; Sequence 58, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREX
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Generic Peptide
US-10-032-221B-58

Query Match 4.1%; Score 10; DB 14; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYSYWL 97
| | | | | | | | | |
Db 1 ASRNDYSYWL 10

RESULT 75

US-10-032-221B-40
; Sequence 40, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 40
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T8-3 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8)
; OTHER INFORMATION: ted for the leucine residue at position 68 of the full-length T8
; OTHER INFORMATION: statin molecule, and serine has been substituted for the cysteine
; OTHER INFORMATION: residues at positions 79 and 85)
US-10-032-221B-40

Query Match 4.1%; Score 10; DB 14; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.027; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRFTTMPFLF 78
| | | | | | | | | |
Db 2 QRFTTMPFLF 11

RESULT 76

US-10-032-221B-42
; Sequence 42, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17

; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: P2 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8)
; OTHER INFORMATION: tatin molecule, and aspartic acid has been substituted for the
; OTHER INFORMATION: steine residues at positions 79 and 85)
US-10-032-221B-42

Query Match 4.1%; Score 10; DB 14; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.027; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRFTTMPFLF 78
| | | | | | | | | |
Db 2 QRFTTMPFLF 11

RESULT 77

US-10-043-487-342
; Sequence 342, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypept
; TITLE OF INVENTION: mammalian polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 342
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Shigella Flexneri
US-10-043-487-342

Query Match 4.1%; Score 10; DB 14; Length 143;

Best Local Similarity 100.0%; Pred. No. 0.12; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
| | | | | | | | | |
Db 27 PAIAIAVHSQ 36

RESULT 78

US-10-206-699-303
; Sequence 303, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 303
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: alpha 2 chain
US-10-206-699-303

Query Match 4.1%; Score 10; DB 14; Length 227;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
Db 111 PAIAIAVHSQ 120

RESULT 79

US-10-032-221B-6
; Sequence 6, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Rachuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-221B-6

Query Match 4.1%; Score 10; DB 14; Length 227;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
Db 111 PAIAIAVHSQ 120

RESULT 80

US-09-925-302-518
; Sequence 518, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302

; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 518
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-518

Query Match 4.1%; Score 10; DB 9; Length 430;

Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
Db 314 PAIAIAVHSQ 323

RESULT 81

US-10-331-496A-27
; Sequence 27, Application US/10331496A
; Publication No. US20030228305A1
; GENERAL INFORMATION:
; APPLICANT: FRANTZ, GRETCHEN
; APPLICANT: HILLAN, KENNETH J.
; APPLICANT: PHILLIPS, HEIDI S.
; APPLICANT: POLAKIS, PAUL
; APPLICANT: SMITH, VICTORIA
; APPLICANT: SPENCER, SUSAN D.
; APPLICANT: WILLIAMS, P. MICKEY
; APPLICANT: WU, THOMAS D.
; APPLICANT: ZHANG, ZHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF TUMOR
; FILE REFERENCE: PS014R1-PCT
; CURRENT APPLICATION NUMBER: US/10/331,496A
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US 60/345,444
; PRIOR FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: US 60/351,885
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/360,066
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: US 60/362,004
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/366,869
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US 60/366,284
; PRIOR FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 60/368,679
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/404,809
; PRIOR FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: US 60/405,645
; PRIOR FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 27
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-331-496A-27

Query Match 4.1%; Score 10; DB 15; Length 459;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
Db 343 PAIAIAVHSQ 352

RESULT 82

US-10-372-683-30
; Sequence 30, Application US/10372683
; Publication No. US20040009171A1
; GENERAL INFORMATION:
; APPLICANT: GERRITSEN, MARY E.
; APPLICANT: PEALE JR., FRANKLIN V.
; APPLICANT: WU, THOMAS D.
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
; FILE REFERENCE: P1928R1P1
; CURRENT APPLICATION NUMBER: US/10/372,683
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 10/271,690
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/344,534
; PRIOR FILING DATE: 2001-10-18
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 30
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-372-683-30

Query Match 4.1%; Score 10; DB 15; Length 459;
Best Local Similarity 100.0%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

QY 128 PAIAIAVHSQ 137
Db 343 PAIAIAVHSQ 352

RESULT 83

US-09-961-403-9
; Sequence 9, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-9

Query Match 4.1%; Score 10; DB 10; Length 1712;
Best Local Similarity 100.0%; Pred. No. 1; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

QY 128 PAIAIAVHSQ 137
Db 1596 PAIAIAVHSQ 1605

RESULT 84

US-10-206-699-9

; Sequence 9, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-9

Query Match 3.7%; Score 9; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.13; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 189 ECHGRGTCN 197
Db 4 ECHGRGTCN 12

RESULT 85

US-10-206-699-76
; Sequence 76, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 76
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-76

Query Match 3.7%; Score 9; DB 14; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 55 RAHQDQLGT 63
Db 6 RAHQDQLGT 14

RESULT 86

US-10-206-699-259
; Sequence 259, Application US/10206699

Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 259
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-259

Query Match 3.7%; Score 9; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92
| | | | | | | | | |
Db 10 VCNFASRND 18

RESULT 87
US-10-206-699-261
; Sequence 261, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 261
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-261

Query Match 3.7%; Score 9; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92
| | | | | | | | | |
Db 10 VCNFASRND 18

RESULT 88
US-10-206-699-265
; Sequence 265, Application US/10206699
; Publication No. US20030100510A1

GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 265
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-265

Query Match 3.7%; Score 9; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92
| | | | | | | | | |
Db 14 VCNFASRND 22

RESULT 89
US-10-206-699-267
; Sequence 267, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 267
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-267

Query Match 3.7%; Score 9; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92
| | | | | | | | | |
Db 14 VCNFASRND 22

RESULT 90
US-10-032-221B-50
; Sequence 50, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Generic Peptide
US-10-032-221B-50

Query Match 3.3%; Score 8; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 9,7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRTTTP 75
Db 1 LQRTTTP 8

RESULT 91
US-10-032-221B-56
; Sequence 56, Application US/10032221B
; Publication No. US2003014481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Generic Peptide
US-10-032-221B-56

Query Match 3.3%; Score 8; DB 14; Length 8;

Best Local Similarity 100.0%; Pred. No. 9,7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 88 ASNDYSY 95
Db 1 ASNDYSY 8
RESULT 92
US-10-270-877-44
; Sequence 44, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII-IV-V
; OTHER INFORMATION: derived peptide
US-10-270-877-44

Query Match 3.3%; Score 8; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 QRAHGQDL 61
Db 1 QRAHGQDL 8

RESULT 93
US-10-270-837-44
; Sequence 44, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII-IV-V
; OTHER INFORMATION: derived peptide
US-10-270-837-44

Query Match 3.3%; Score 8; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 QRAHGQDL 61

Db 1 QRAHGQDL 8

RESULT 94

US-10-206-699-253
; Sequence 253, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 253
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-253

Query Match 3.3%; Score 8; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TWPFLFCN 80

Db 3 TWPFLFCN 10

RESULT 95

US-10-270-877-43
; Sequence 43, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII derived
US-10-270-877-43

Query Match 3.3%; Score 8; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61

Db 1 QRAHGQDL 8

RESULT 96

US-10-270-837-43
; Sequence 43, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII derived
US-10-270-837-43

Query Match 3.3%; Score 8; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61

Db 1 QRAHGQDL 8

RESULT 97

US-10-206-699-288
; Sequence 288, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 288
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-288

Query Match 3.3%; Score 8; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TWPFLFCN 80

Db 6 TWPFLFCN 13

RESULT 98

US-10-206-699-307
; Sequence 307, Application US/10206699

Publication No. US20030100510A1
GENERAL INFORMATION:
APPLICANT: Sundaramoorthy, M.
TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
FILE REFERENCE: MBHB 01-1017
CURRENT APPLICATION NUMBER: US/10/206,699
CURRENT FILING DATE: 2002-07-26
PRIOR APPLICATION NUMBER: US 60/308,523
PRIOR FILING DATE: 2001-07-27
PRIOR APPLICATION NUMBER: US 60/351,289
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: US 60/366,854
PRIOR FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: US 60/385,362
PRIOR FILING DATE: 2002-06-03
NUMBER OF SEQ ID NOS: 307
SOFTWARE: Patentin version 3.1
SEQ ID NO 307
LENGTH: 228
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: alpha 6 chain
US-10-206-699-307

Query Match 3.3%; Score 8; DB 14; Length 228;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
Db 219 SRCQVCWK 226
|||||

RESULT 99

US-10-369-493-3337
Sequence 3337, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 3337
LENGTH: 937
TYPE: PRT
ORGANISM: Neurospora crassa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(937)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3337

Query Match 3.3%; Score 8; DB 15; Length 937;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 IMFTSAGS 165
Db 576 IMFTSAGS 583
|||||

RESULT 100
US-10-032-221B-49
Sequence 49, Application US/10032221B
Publication No. US20030144481A1
GENERAL INFORMATION:
APPLICANT: Kalluri, Raghuram
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/335,224
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/089,689
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentin version 3.1
SEQ ID NO 49
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Generic peptide
US-10-032-221B-49

Query Match 2.9%; Score 7; DB 14; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.7e-05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRFTTMP 75
Db 1 QRFTTMP 7
|||||

RESULT 101

US-10-032-221B-55
Sequence 55, Application US/10032221B
Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
CURRENT APPLICATION NUMBER: US/10/032,221B
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: PCT/US01/00565
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: US 09/625,191
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 09/543,371
PRIOR FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: US 09/479,118
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/335,224
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/126,175
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/089,689
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentin version 3.1
SEQ ID NO 55
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Generic Peptide
US-10-032-221B-55

Query Match 2.9%; Score 7; DB 14; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYS 94
|||||
Db 1 ASRNDYS 7

RESULT 102

US-10-032-221B-51
; Sequence 51, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Generic Peptide
US-10-032-221B-51

Query Match 2.9%; Score 7; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRETTMP 75
|||||
Db 2 QRETTMP 8

RESULT 103

US-10-206-699-26
; Sequence 26, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-26

Query Match 2.9%; Score 7; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMFFLFC 79
|||||
Db 8 TMFFLFC 14

RESULT 104

US-10-206-699-92
; Sequence 92, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 92
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-92

Query Match 2.9%; Score 7; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GQALASP 175
|||||
Db 9 GQALASP 15

RESULT 105

US-10-206-699-230
; Sequence 230, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22

; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 230
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-230

Query Match 2.9%; Score 7; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 193 RGTCTNY 199
Db 1 RGTCTNY 7

RESULT 106

US-10-206-699-232
; Sequence 232, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT FILING DATE: 2002-07-26
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 232
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-232

Query Match 2.9%; Score 7; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 193 RGTCTNY 199
Db 1 RGTCTNY 7

RESULT 107

US-09-864-761-37448
; Sequence 37448, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37448
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035425.11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: P29400, EVALUATE 9.00e-24
; OTHER INFORMATION: EST_HUMAN HIT: W07655.1, EVALUATE 6.00e-23
US-09-864-761-37448

Query Match 2.9%; Score 7; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 TRHSQTT 29
Db 32 TRHSQTT 38

RESULT 108

US-09-864-761-47938
; Sequence 47938, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

Query Match 2.9%; Score 7; DB 9; Length 87;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GKRGSQ 10
Db 24 GKRGSQ 30

RESULT 112

US-10-242-515-1521
; Sequence 1521, Application US/10242515
; Publication No. US2004009488A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005C1
; CURRENT APPLICATION NUMBER: US/10/242,515
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1521
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-242-515-1521

Query Match 2.9%; Score 7; DB 15; Length 87;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GKRGSQ 10
Db 24 GKRGSQ 30

RESULT 113

US-10-425-114-54455
; Sequence 54455, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 54455
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17114F11_FLI.pep
US-10-425-114-54455

Query Match

Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 HGQDLGT 63
Db 102 HGQDLGT 108

RESULT 114

US-09-461-580A-18
; Sequence 18, Application US/09461580A
; Publication No. US20030207325A1
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichiro
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-461-580A-18

Query Match 2.9%; Score 7; DB 11; Length 117;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 KPIPTV 224
Db 110 KPIPTV 116

RESULT 115

US-09-864-761-38021
; Sequence 38021, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 38021
LENGTH: 142
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031177.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.99
OTHER INFORMATION: SWISSPROT HIT: Q14031, EVALUATE 3.00e-55
OTHER INFORMATION: EST_HUMAN HIT: A0142039.1, EVALUATE 2.00e-42
US-09-864-761-38021

Query Match 2.9%; Score 7; DB 9; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AIAVHSQ 137
DB 82 AIAVHSQ 88

RESULT 116
US-10-094-749-3238
Sequence 3238, Application US/10094749
Publication No. US20030219741A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHIKO
APPLICANT: YOSHIKAWA, TSUTOMU

APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3238
LENGTH: 150
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-3238

Query Match 2.9%; Score 7; DB 15; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 117
US-10-425-114-47774
Sequence 47774, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovacic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 47774
LENGTH: 174
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3960-011-H7_FLI.pgp
US-10-425-114-47774

Query Match 2.9%; Score 7; DB 12; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.3e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
DB 68 ALASPGS 74

RESULT 118
US-09-727-855B-7
Sequence 7, Application US/09727855B
Patent No. US20020168703A1
GENERAL INFORMATION:
APPLICANT: FOSHINO, Tatsuo
APPLICANT: OJIMA, Kazuyuki
APPLICANT: SETOGUCHI, Yutaka
TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF CAROTENOIDS AND BIOLOGICALLY USE
TITLE OF INVENTION: MATERIALS THEREOF
FILE REFERENCE: C38435/111694

; CURRENT APPLICATION NUMBER: US/09/727,855B
; CURRENT FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Phaffia rhodozyma
US-09-727-855B-7

Query Match 2.9%; Score 7; DB 9; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYIS 120
Db 15 ALEPYIS 21

RESULT 119

; Sequence 59362, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yinhua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59362
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Oryza sativa nipponbare
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-OSLELIB3474052F08_FLI.pep
US-10-425-114-59362

Query Match 2.9%; Score 7; DB 12; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYIS 120
Db 63 ALEPYIS 69

RESULT 120

; Sequence 96, Application US/10267682
; Publication No. US2004003235A1
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TRANSMISSION
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; COUNTRY: USA
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/267,748
; FILING DATE: 08-Oct-2002

; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/267,682
; FILING DATE: 08-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/484,223A
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cortuzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 96:
US-10-267-682-96

Query Match 2.9%; Score 7; DB 12; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGIGQA 171
Db 29 SEGIGQA 35

RESULT 121

; Sequence 96, Application US/10267748
; Publication No. US20040052820A1
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TRANSMISSION
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/267,748
; FILING DATE: 08-Oct-2002

Qy 165 SEGIGQA 171
Db 29 SEGIGQA 35

US-10-267-748-96
; Sequence 96, Application US/10267748
; Publication No. US20040052820A1
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TRANSMISSION
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/267,748
; FILING DATE: 08-Oct-2002

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/484,223A
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 96:
US-10-267-748-96

Query Match 2.9%; Score 7; DB 12; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGOA 171
Db 29 SEGTGOA 35
|||||

RESULT 122
US-10-206-699-305
Sequence 305, Application US/10206699
Publication No. US20030100510A1
GENERAL INFORMATION:
APPLICANT: Sundaramoorthy, M.
APPLICANT: Hudson, B.
TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
FILE REFERENCE: MBHB 01-1017
CURRENT APPLICATION NUMBER: US/10/206,699
CURRENT FILING DATE: 2002-07-26
PRIOR APPLICATION NUMBER: US 60/308,523
PRIOR FILING DATE: 2001-07-27
PRIOR APPLICATION NUMBER: US 60/351,289
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: US 60/366,854
PRIOR FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: US 60/385,362
PRIOR FILING DATE: 2002-06-03
NUMBER OF SEQ ID NOS: 307
SOFTWARE: Patent in version 3.1
SEQ ID NO 305
LENGTH: 231
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: alpha 4 chain
US-10-206-699-305

Query Match 2.9%; Score 7; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPQSCLE 180
Db 159 SPQSCLE 165
|||||

RESULT 123
US-10-230-331-28
Sequence 28, Application US/10230331

Publication No. US20030157513A1
GENERAL INFORMATION:
APPLICANT: RAJASEKHARAN, Ram
TITLE OF INVENTION: A NOVEL TRIACYLGLYCEROL BIOSYNTHESIS IN THE CYTOSOL OF EUKARYOTE
FILE REFERENCE: 110522
CURRENT APPLICATION NUMBER: US/10/230,331
CURRENT FILING DATE: 2002-08-29
PRIOR APPLICATION NUMBER: US 60/315,757
PRIOR FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patent in version 3.2
SEQ ID NO 28
LENGTH: 233
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-230-331-28

Query Match 2.9%; Score 7; DB 14; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYIS 120
Db 39 ALEPYIS 45
|||||

RESULT 124
US-10-282-122A-72960
Sequence 72960, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 72960
LENGTH: 247
TYPE: PRT

; ORGANISM: Salmonella paratyphi A
US-10-282-122A-72960

Query Match 2.9%; Score 7; DB 12; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212
Db 157 WLASLNP 163

RESULT 125

US-10-282-122A-76032
; Sequence 76032, Application US/10282122A
; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-03-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-03-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO: 76032

; TYPE: PRT

; ORGANISM: Salmonella typhi

US-10-282-122A-76032

Query Match 2.9%; Score 7; DB 12; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212
Db 157 WLASLNP 163

RESULT 126

US-10-296-115-808

; Sequence 808, Application US/10296115
; Publication No. US20040053248A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT

; CURRENT APPLICATION NUMBER: US/10/296,115

; CURRENT FILING DATE: 2002-11-18

; PRIOR APPLICATION NUMBER: US09/488,725

; PRIOR FILING DATE: 2000-01-21

; PRIOR APPLICATION NUMBER: US09/552,317

; PRIOR FILING DATE: 2000-04-25

; NUMBER OF SEQ ID NOS: 1478

; SEQ ID NO: 808

; LENGTH: 251

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)..(251)

; OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3

US-10-296-115-808

Query Match 2.9%; Score 7; DB 12; Length 251;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177

Db 185 ALASPGS 191

RESULT 127

US-10-424-599-235900

; Sequence 235900, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO: 235900

; LENGTH: 300

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_55045C.1.pgp

US-10-424-599-235900

Query Match 2.9%; Score 7; DB 12; Length 300;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 LEEFRAS 185

Db 147 LEEFRAS 153

RESULT 128

US-09-964-956-71

; Sequence 71, Application US/09964956

; Publication No. US20040043926A1

; GENERAL INFORMATION:

; APPLICANT: Gerlach, Valerie L

; APPLICANT: MacDougall, John R

; APPLICANT: Smithson, Glennda

; APPLICANT: Millet, Isabelle

```

; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lopley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: No. US2004043926A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/238,399
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/238,396
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/276,667
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/294,823
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/304,868
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: CNH domain
; OTHER INFORMATION: Consensus Sequence
US-09-964-956-71

Query Match          2.9%; Score 7; DB 12; Length 304;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      229 LEKISR 235
DB      36 LEKISR 42

RESULT 129
US-10-425-114-59795
; Sequence 59795, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59795
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-095-Cl1_FLI.pep
US-10-425-114-59795

Query Match          2.9%; Score 7; DB 12; Length 311;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      114 ALEPYIS 120
DB      128 ALEPYIS 134

RESULT 130
US-10-425-114-71839
; Sequence 71839, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71839
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB7316SF09_FLI.pep
US-10-425-114-71839

Query Match          2.9%; Score 7; DB 12; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 KRGDGSGS 11
DB      207 KRGDGSGS 213

RESULT 131
US-10-425-114-50147
; Sequence 50147, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B

```

```

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59795
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-095-Cl1_FLI.pep
US-10-425-114-59795

Query Match          2.9%; Score 7; DB 12; Length 311;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      114 ALEPYIS 120
DB      128 ALEPYIS 134

RESULT 130
US-10-425-114-71839
; Sequence 71839, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B

```

```
;
;
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 50147
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700901538_FLI.pap
; US-10-425-114-50147

Query Match      2.9%; Score 7; DB 12; Length 317;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 IPSTVKA 226
DB 288 IPSTVKA 294

RESULT 132
US-10-424-599-201843
; Sequence 201843, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 201843
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(328)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_2428C.1.pap
; US-10-424-599-201843

Query Match      2.9%; Score 7; DB 12; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 IPSTVKA 226
DB 299 IPSTVKA 305

RESULT 133
US-09-918-568-58
; Sequence 58, Application US/09918568
; Patent No. US20020054882A1
; GENERAL INFORMATION:
; APPLICANT: Yoshitobu OKUNO et al.
; TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING
; ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/918,568
; FILING DATE: 02-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/004,422
; FILING DATE: January 8, 1998
; APPLICATION NUMBER: 08/443,862
; FILING DATE: May 22, 1995
; APPLICATION NUMBER: 08/229,781
; FILING DATE: April 19, 1994
; APPLICATION NUMBER: 08/054,016
; FILING DATE: April 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: <Unknown>
; ANTI-SENSE: <Unknown>
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; ORGANISM: <Unknown>
; STRAIN: <Unknown>
; INDIVIDUAL ISOLATE: <Unknown>
; DEVELOPMENTAL STAGE: <Unknown>
; HAPLOTYPE: <Unknown>
; TISSUE TYPE: <Unknown>
; CELL TYPE: <Unknown>
; CELL LINE: <Unknown>
; ORGANELLE: <Unknown>
; IMMEDIATE SOURCE:
; LIBRARY: <Unknown>
; CLONE: <Unknown>
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: <Unknown>
; MAP POSITION: <Unknown>
; UNITS: <Unknown>
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
; US-09-918-568-58

Query Match      2.9%; Score 7; DB 9; Length 347;
```

```
Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 165 SEGTGQA 171
Db 155 SEGTGQA 161

RESULT 134
US-09-860-351-2
; Sequence 2, Application US/09860351
; Patent No. US20020077463A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 16105, A NOVEL PROTEIN HUMAN PHOSPHATASE
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 38155-20013.00
; CURRENT APPLICATION NUMBER: US/09/860,351
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/205,260
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-860-351-2

Query Match 2.9%; Score 7; DB 9; Length 352;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 135
US-10-353-690-44
; Sequence 44, Application US/10353690
; Publication No. US20030215840A1
; GENERAL INFORMATION:
; APPLICANT: Logan, Thomas Joseph
; APPLICANT: Chun, Miyoung
; APPLICANT: Galvin, Katherine M.
; APPLICANT: Healy, Aileen
; APPLICANT: Acton, Susan L.
; APPLICANT: Donoghue, Mary
; APPLICANT: Stagliano, Nancy
; APPLICANT: Perodini, Jacqueline
; APPLICANT: Rodrigue-Way, Amelie
; TITLE OF INVENTION: Methods and compositions for treating
; TITLE OF INVENTION: cardiovascular disease using 1682, 6169, 6193, 7771, 14395,
; TITLE OF INVENTION: 29002, 33216, 43726, 62922, 66156, 32427, 2402, 7747, 1720,
; TITLE OF INVENTION: 9151, 60491, 1371, 7077, 33207, 1419, 18036, 16105, 38650,
; TITLE OF INVENTION: 14245, 58848, 1870, 25856, 32394, 3484, 345, 9252, 9135,
; TITLE OF INVENTION: 10532, 18610, 8165, 2448, 2445, 64624, 84237, 8912, 2868,
; TITLE OF INVENTION: 283, 2554, 9464, 17799, 26686, 43848, 32135, 12208, 2914,
; TITLE OF INVENTION: 51130, 19489, 21833, 2917, 59590, 15992, 2094, 2252, 3474,
; TITLE OF INVENTION: 9792, 15400, 1452 or 6585 molecules
; FILE REFERENCE: ME102-018PIRONNIM
; CURRENT APPLICATION NUMBER: US/10/353,690
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/353,224
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/364,529
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 60/373,861
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/376,287
; PRIOR FILING DATE: 2002-04-29
```

```
; PRIOR APPLICATION NUMBER: 60/388,080
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: 60/390,971
; PRIOR FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/394,130
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: 60/394,797
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: 60/404,904
; PRIOR FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/405,450
; PRIOR FILING DATE: 2002-08-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-353-690-44

Query Match 2.9%; Score 7; DB 15; Length 352;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 136
US-09-973-963-4
; Sequence 4, Application US/09973963
; Patent No. US20020106676A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,963
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: US 60/304,775
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-973-963-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 137
US-09-973-064-4
; Sequence 4, Application US/09973064
; Patent No. US20020106773A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
```

; TITLE OF INVENTION: Diseases
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,064
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-973-064-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 138

US-09-973-077-4
; Sequence 4, Application US/09973077
; Patent No. US20020114799A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,077
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-973-077-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 139

US-09-973-063-4
; Sequence 4, Application US/09973063
; Patent No. US20020115119A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,063
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4

; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-973-063-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 140

US-09-973-964-4
; Sequence 4, Application US/09973964
; Patent No. US20020115606A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,964
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: US 60/304,775
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-973-964-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 141

US-09-975-072-4
; Sequence 4, Application US/09975072
; Patent No. US20020115607A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/975,072
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-975-072-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 142

US-09-972-038-4
; Sequence 4, Application US/09972038
; Patent No. US20020119155A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/972,038
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-038-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 143

US-09-972-757-4
; Sequence 4, Application US/09972757
; Patent No. US20020119927A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/972,757
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-757-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 144

US-09-973-965-4

; Sequence 4, Application US/09973965
; Patent No. US20020124273A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,965
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: US 60/304,775
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-973-965-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 145

US-09-973-941-4
; Sequence 4, Application US/09973941
; Patent No. US20020164655A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,941
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: US 60/304,775
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-973-941-4

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 146

US-09-986-992-2
; Sequence 2, Application US/09986992
; Publication No. US20030027308A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID

```
; APPLICANT: MANNING, GERARD
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN PHOSPHATASES IDENTIFIED FROM
; FILE REFERENCE: 038602/1277
; CURRENT APPLICATION NUMBER: US/09/986,992
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 60/246,974
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/208,291
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-992-2

Query Match      2.9%; Score 7; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 147
US-09-971-782-4
; Sequence 4, Application US/09971782
; Publication No. US20030186317A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/971,782
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-971-782-4

Query Match      2.9%; Score 7; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 148
US-09-971-782-4
; Sequence 4, Application US/09971782
; Publication No. US20030186317A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/971,782
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/240,790
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-971-782-4

Query Match      2.9%; Score 7; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 149
US-10-094-749-1699
; Sequence 1699, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
```

```
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NACHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1699
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1699

Query Match      2.9%; Score 7; DB 15; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 149
US-10-311-764-1
; Sequence 1, Application US/10311764
; Publication No. US20040023245A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; AU-YOUNG, Janice K.
; APPLICANT: BAUGHN, Mariah R.; DING, Li
; APPLICANT: ELLIOTT, Vicki S.; GANDHI, Ameena R.
; APPLICANT: GRIFFIN, Jennifer A.; HARALIA, April J.A.
; APPLICANT: KEARNEY, Liam; LEE, Ernestine A.
; APPLICANT: LU, Yan; NGUYEN, Daniel B.
; APPLICANT: ARVIZU, Chandra S.; RAMKUMAR, Jayalaxmi
; APPLICANT: REDDY, Rooda M.; SANJANWALA, Madhusudan M.
; APPLICANT: STEWART, Elizabeth A.; TANG, Y. Tom
; APPLICANT: THORNTON, Michael B.; TRIBOULEY, Catherine M.
; APPLICANT: CHAWLA, Narinder K.; YANG, Junming
; APPLICANT: YAO, Monique G.; YUE, Henry
; TITLE OF INVENTION: PROTEIN PHOSPHATASES
; FILE REFERENCE: PI-0126 USN
; CURRENT APPLICATION NUMBER: US/10/311,764
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/US01/19442
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/212,447
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/213,746
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: US 60/215,210
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/216,529
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/218,080
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/220,117
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 372
```



```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023245A1 8124196CD1
US-10-311-764-1
```

```
Query Match      2.9%; Score 7; DB 16; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 9 SGSPATW 15
Db 55 SGSPATW 61
```

RESULT 150

```
US-09-925-300-1655
; Sequence 1655, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: FA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 1655
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (290)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (325)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1655
```

```
Query Match      2.9%; Score 7; DB 9; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 9 SGSPATW 15
Db 56 SGSPATW 62
```

Search completed: April 5, 2004, 07:44:40
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:36:14 ; Search time 22 Seconds
(without alignments)
572.579 Million cell updates/sec

Title: US-10-032-221B-10
Perfect score: 244
Sequence: 1 GLKRGDSGSPATWTRGF.....KAGLEKLIISRCQVCMKKRH 244

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/2/iaa/5A-COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B-COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A-COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B-COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS-COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 163 | 66.8 | 268 | 4 | US-09-589-927-6 |
| 2 | 163 | 66.8 | 268 | 4 | US-09-589-927-6 |
| 3 | 163 | 66.8 | 268 | 4 | US-09-589-927-6 |
| 4 | 159 | 65.2 | 211 | 4 | US-09-512-563C-46 |
| 5 | 141 | 57.8 | 218 | 2 | US-08-399-889-25 |
| 6 | 141 | 57.8 | 218 | 3 | US-09-167-364-25 |
| 7 | 141 | 57.8 | 218 | 3 | US-09-439-897-4 |
| 8 | 61 | 25.0 | 68 | 4 | US-09-512-563C-50 |
| 9 | 61 | 25.0 | 72 | 4 | US-09-512-563C-48 |
| 10 | 39 | 16.0 | 471 | 2 | US-08-399-889-24 |
| 11 | 39 | 16.0 | 471 | 3 | US-09-167-364-24 |
| 12 | 39 | 16.0 | 471 | 3 | US-09-439-897-2 |
| 13 | 37 | 15.2 | 72 | 4 | US-09-512-563C-61 |
| 14 | 36 | 14.8 | 36 | 3 | US-09-439-897-63 |
| 15 | 36 | 14.8 | 36 | 3 | US-09-439-897-65 |
| 16 | 26 | 10.7 | 26 | 3 | US-09-512-563C-26 |
| 17 | 21 | 8.6 | 21 | 4 | US-09-589-927-2 |
| 18 | 17 | 7.0 | 260 | 4 | US-09-277-665-2 |
| 19 | 17 | 7.0 | 260 | 4 | US-09-589-987-2 |
| 20 | 17 | 7.0 | 260 | 4 | US-09-589-987-2 |
| 21 | 17 | 7.0 | 264 | 4 | US-09-589-927-10 |
| 22 | 17 | 7.0 | 264 | 4 | US-09-277-665-10 |
| 23 | 17 | 7.0 | 264 | 4 | US-09-589-987-10 |
| 24 | 15 | 6.1 | 15 | 3 | US-09-439-897-53 |
| 25 | 15 | 6.1 | 15 | 3 | US-09-439-897-57 |
| 26 | 15 | 6.1 | 15 | 3 | US-09-439-897-59 |
| 27 | 15 | 6.1 | 15 | 3 | US-09-439-897-61 |

| | | | | | | | | | | | | | |
|-----|---|-----|------|---|----------------------|--------------------|-----|---|-----|-----|---|----------------------|--------------------|
| 101 | 7 | 2.9 | 571 | 2 | US-08-453-848-21 | Sequence 21, Appl | 174 | 6 | 2.5 | 111 | 1 | US-08-466-886-38 | Sequence 38, Appl |
| 102 | 7 | 2.9 | 571 | 3 | US-09-169-027-15 | Sequence 15, Appl | 175 | 6 | 2.5 | 111 | 3 | US-08-469-617-38 | Sequence 38, Appl |
| 103 | 7 | 2.9 | 571 | 3 | US-09-169-027-21 | Sequence 21, Appl | 176 | 6 | 2.5 | 121 | 4 | US-09-621-976-7342 | Sequence 7342, App |
| 104 | 7 | 2.9 | 599 | 1 | US-08-463-163-3 | Sequence 3, Appl | 177 | 6 | 2.5 | 122 | 4 | US-09-325-932A-157 | Sequence 157, App |
| 105 | 7 | 2.9 | 599 | 1 | US-08-405-615-1 | Sequence 1, Appl | 178 | 6 | 2.5 | 126 | 4 | US-09-134-000C-5187 | Sequence 5187, Ap |
| 106 | 7 | 2.9 | 613 | 2 | US-08-461-234-1 | Sequence 1, Appl | 179 | 6 | 2.5 | 128 | 4 | US-09-252-991A-21538 | Sequence 21538, A |
| 107 | 7 | 2.9 | 613 | 2 | US-08-463-480-1 | Sequence 1, Appl | 180 | 6 | 2.5 | 129 | 4 | US-09-543-681A-6111 | Sequence 6111, Ap |
| 108 | 7 | 2.9 | 613 | 4 | US-09-479-479-2 | Sequence 2, Appl | 181 | 6 | 2.5 | 131 | 4 | US-09-553-498-4 | Sequence 4, Appl |
| 109 | 7 | 2.9 | 613 | 4 | US-09-297-851-2 | Sequence 2, Appl | 182 | 6 | 2.5 | 131 | 4 | US-09-618-869-4 | Sequence 4, Appl |
| 110 | 7 | 2.9 | 614 | 1 | US-08-225-224-1 | Sequence 1, Appl | 183 | 6 | 2.5 | 135 | 3 | US-09-393-395-9 | Sequence 9, Appl |
| 111 | 7 | 2.9 | 614 | 3 | US-08-722-258-1 | Sequence 1, Appl | 184 | 6 | 2.5 | 135 | 3 | US-09-668-648-9 | Sequence 9, Appl |
| 112 | 7 | 2.9 | 614 | 5 | PCT-US95-04468-1 | Sequence 1, Appl | 185 | 6 | 2.5 | 140 | 4 | US-09-724-138-44 | Sequence 44, Appl |
| 113 | 7 | 2.9 | 618 | 4 | US-09-370-516-4 | Sequence 4, Appl | 186 | 6 | 2.5 | 144 | 4 | US-09-252-991A-20518 | Sequence 20518, A |
| 114 | 7 | 2.9 | 622 | 3 | US-08-356-786-16 | Sequence 16, Appl | 187 | 6 | 2.5 | 146 | 4 | US-08-352-991A-23680 | Sequence 23680, A |
| 115 | 7 | 2.9 | 632 | 3 | US-09-546-992-2 | Sequence 2, Appl | 188 | 6 | 2.5 | 150 | 4 | US-09-636-215-707 | Sequence 707, App |
| 116 | 7 | 2.9 | 637 | 1 | US-08-335-838-14 | Sequence 14, Appl | 189 | 6 | 2.5 | 150 | 4 | US-09-685-166A-707 | Sequence 707, App |
| 117 | 7 | 2.9 | 637 | 1 | US-08-235-838-16 | Sequence 16, Appl | 190 | 6 | 2.5 | 151 | 3 | US-09-188-930-276 | Sequence 276, App |
| 118 | 7 | 2.9 | 637 | 2 | US-08-465-473B-14 | Sequence 14, Appl | 191 | 6 | 2.5 | 151 | 4 | US-09-312-283C-276 | Sequence 276, App |
| 119 | 7 | 2.9 | 637 | 2 | US-08-465-473B-16 | Sequence 16, Appl | 192 | 6 | 2.5 | 152 | 4 | US-09-621-976-4169 | Sequence 4169, Ap |
| 120 | 7 | 2.9 | 638 | 3 | US-09-047-148-2 | Sequence 2, Appl | 193 | 6 | 2.5 | 153 | 3 | US-08-828-741B-11 | Sequence 11, Appl |
| 121 | 7 | 2.9 | 658 | 4 | US-08-252-991A-13879 | Sequence 13879, A | 194 | 6 | 2.5 | 153 | 4 | US-08-925-433-4 | Sequence 4, Appl |
| 122 | 7 | 2.9 | 676 | 2 | US-08-398-590A-40 | Sequence 40, Appl | 195 | 6 | 2.5 | 155 | 4 | US-09-160-567-11 | Sequence 11, Appl |
| 123 | 7 | 2.9 | 676 | 3 | US-08-894-997-40 | Sequence 40, Appl | 196 | 6 | 2.5 | 155 | 4 | US-09-553-132-4 | Sequence 4, Appl |
| 124 | 7 | 2.9 | 767 | 4 | US-09-252-991A-13361 | Sequence 13361, A | 197 | 6 | 2.5 | 155 | 4 | US-09-710-299-11 | Sequence 11, Appl |
| 125 | 7 | 2.9 | 1694 | 1 | US-08-494-168-2 | Sequence 2, Appl | 198 | 6 | 2.5 | 155 | 4 | US-09-509-031-11 | Sequence 11, Appl |
| 126 | 7 | 2.9 | 3562 | 2 | US-09-679-279-14 | Sequence 14, Appl | 199 | 6 | 2.5 | 166 | 2 | US-08-729-103-4 | Sequence 4, Appl |
| 127 | 6 | 2.5 | 6 | 3 | US-09-439-897-51 | Sequence 51, Appl | 200 | 6 | 2.5 | 168 | 2 | US-08-702-105A-29 | Sequence 29, Appl |
| 128 | 6 | 2.5 | 7 | 4 | US-10-080-505-58 | Sequence 58, Appl | | | | | | | |
| 129 | 6 | 2.5 | 8 | 3 | US-08-296-791-7 | Sequence 7, Appl | | | | | | | |
| 130 | 6 | 2.5 | 8 | 3 | US-08-296-791-8 | Sequence 8, Appl | | | | | | | |
| 131 | 6 | 2.5 | 8 | 4 | US-09-839-996-7 | Sequence 7, Appl | | | | | | | |
| 132 | 6 | 2.5 | 8 | 4 | US-09-839-996-8 | Sequence 8, Appl | | | | | | | |
| 133 | 6 | 2.5 | 8 | 4 | US-10-080-505-53 | Sequence 53, Appl | | | | | | | |
| 134 | 6 | 2.5 | 8 | 4 | US-10-080-505-54 | Sequence 54, Appl | | | | | | | |
| 135 | 6 | 2.5 | 8 | 5 | PCT-US95-10661A-7 | Sequence 7, Appl | | | | | | | |
| 136 | 6 | 2.5 | 8 | 5 | PCT-US95-10661A-8 | Sequence 8, Appl | | | | | | | |
| 137 | 6 | 2.5 | 9 | 4 | US-09-322-624-4 | Sequence 4, Appl | | | | | | | |
| 138 | 6 | 2.5 | 11 | 4 | US-10-080-505-20 | Sequence 20, Appl | | | | | | | |
| 139 | 6 | 2.5 | 12 | 1 | US-08-260-582-40 | Sequence 40, Appl | | | | | | | |
| 140 | 6 | 2.5 | 12 | 5 | PCT-US95-05471-40 | Sequence 40, Appl | | | | | | | |
| 141 | 6 | 2.5 | 15 | 3 | US-09-439-897-58 | Sequence 58, Appl | | | | | | | |
| 142 | 6 | 2.5 | 16 | 4 | US-09-106-568B-149 | Sequence 149, App | | | | | | | |
| 143 | 6 | 2.5 | 19 | 4 | US-08-716-249-1 | Sequence 1, Appl | | | | | | | |
| 144 | 6 | 2.5 | 20 | 4 | US-08-716-249-9 | Sequence 9, Appl | | | | | | | |
| 145 | 6 | 2.5 | 21 | 2 | US-08-449-287-20 | Sequence 20, Appl | | | | | | | |
| 146 | 6 | 2.5 | 21 | 2 | US-09-003-081-8 | Sequence 8, Appl | | | | | | | |
| 147 | 6 | 2.5 | 21 | 3 | US-08-648-262-8 | Sequence 8, Appl | | | | | | | |
| 148 | 6 | 2.5 | 21 | 3 | US-08-648-263-8 | Sequence 8, Appl | | | | | | | |
| 149 | 6 | 2.5 | 21 | 4 | US-08-840-713-48 | Sequence 48, Appl | | | | | | | |
| 150 | 6 | 2.5 | 21 | 4 | US-09-904-196B-3 | Sequence 3, Appl | | | | | | | |
| 151 | 6 | 2.5 | 21 | 4 | US-09-230-233A-2 | Sequence 2, Appl | | | | | | | |
| 152 | 6 | 2.5 | 21 | 4 | US-09-760-008A-3 | Sequence 3, Appl | | | | | | | |
| 153 | 6 | 2.5 | 21 | 4 | US-09-393-171-19 | Sequence 19, Appl | | | | | | | |
| 154 | 6 | 2.5 | 22 | 1 | US-08-318-193-46 | Sequence 46, Appl | | | | | | | |
| 155 | 6 | 2.5 | 23 | 1 | US-08-215-138-1 | Sequence 1, Appl | | | | | | | |
| 156 | 6 | 2.5 | 23 | 1 | US-08-407-344-1 | Sequence 1, Appl | | | | | | | |
| 157 | 6 | 2.5 | 26 | 3 | US-08-866-354A-11 | Sequence 11, Appl | | | | | | | |
| 158 | 6 | 2.5 | 34 | 1 | US-08-190-029A-6 | Sequence 6, Appl | | | | | | | |
| 159 | 6 | 2.5 | 34 | 2 | US-08-462-695-6 | Sequence 6, Appl | | | | | | | |
| 160 | 6 | 2.5 | 35 | 1 | US-08-318-193-47 | Sequence 47, Appl | | | | | | | |
| 161 | 6 | 2.5 | 62 | 4 | US-09-621-976-5709 | Sequence 5709, Ap | | | | | | | |
| 162 | 6 | 2.5 | 62 | 4 | US-09-621-976-5715 | Sequence 5715, Ap | | | | | | | |
| 163 | 6 | 2.5 | 68 | 4 | US-09-134-001C-3669 | Sequence 3669, Ap | | | | | | | |
| 164 | 6 | 2.5 | 68 | 4 | US-09-134-001C-3794 | Sequence 3794, Ap | | | | | | | |
| 165 | 6 | 2.5 | 68 | 4 | US-09-134-001C-4489 | Sequence 4489, Ap | | | | | | | |
| 166 | 6 | 2.5 | 68 | 4 | US-09-134-001C-5041 | Sequence 5041, Ap | | | | | | | |
| 167 | 6 | 2.5 | 72 | 4 | US-09-134-001C-5281 | Sequence 5281, Ap | | | | | | | |
| 168 | 6 | 2.5 | 78 | 6 | 5187153-25 | Patent No. 5187153 | | | | | | | |
| 169 | 6 | 2.5 | 78 | 6 | 5223482-27 | Patent No. 5223482 | | | | | | | |
| 170 | 6 | 2.5 | 92 | 6 | 5171841-1 | Patent No. 5171841 | | | | | | | |
| 171 | 6 | 2.5 | 96 | 4 | US-09-390-134B-53 | Sequence 53, Appl | | | | | | | |
| 172 | 6 | 2.5 | 96 | 4 | US-09-540-236-2315 | Sequence 2315, Ap | | | | | | | |
| 173 | 5 | 2.5 | 108 | 4 | US-09-543-661A-6983 | Sequence 6983, Ap | | | | | | | |

ALIGNMENTS

RESULT 1

US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PPT
; ORGANISM: Human
US-09-589-927-6

Query Match 66.8%; Score 163; DB 4; Length 268;

Best Local Similarity 100.0%; Pred. NO. 2.7e-159;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRGSQSPATWTTGTFVTRHSQTATPSCPECTVPLYSGLFVQGNQRAHQDGLGTL 64

DB 29 KRGSQSPATWTTGTFVTRHSQTATPSCPECTVPLYSGLFVQGNQRAHQDGLGTL 88

QY 65 GSCILQRTFTTMEFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGTRALEPYSRCTV 124

DB 89 GSCILQRTFTTMEFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGTRALEPYSRCTV 148

QY 125 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSLMFTSAGSEG 167

DB 149 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSLMFTSAGSEG 191

RESULT 2

US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:

APPLICANT: University of Kansas Medical Center
TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
FILE REFERENCE: 94525-1
CURRENT APPLICATION NUMBER: US/09/277,665
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 6
LENGTH: 268
TYPE: PRT
ORGANISM: Human
US-09-277-665-6

Query Match 66.8%; Score 163; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 2.7e-159;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRGDGSPATWTRGVFTRHSQTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 64
DB 29 KRGDGSPATWTRGVFTRHSQTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 88
QY 65 GSCLOQRTTTPFLFCNNVNDVCFASNDYSYWLSTPALMPMNPITGRALEPYISRCTV 124
DB 89 GSCLOQRTTTPFLFCNNVNDVCFASNDYSYWLSTPALMPMNPITGRALEPYISRCTV 148
QY 125 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMTSAGSEG 167
DB 149 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMTSAGSEG 191

RESULT 3
US-09-589-987-6
Sequence 6, Application US/095898987
Patent No. 6498140
GENERAL INFORMATION:
APPLICANT: University of Kansas Medical Center
TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
FILE REFERENCE: 945251
CURRENT APPLICATION NUMBER: US/09/589,987
CURRENT FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 6
LENGTH: 268
TYPE: PRT
ORGANISM: Human
US-09-589-987-6

Query Match 66.8%; Score 163; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 2.7e-159;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRGDGSPATWTRGVFTRHSQTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 64
DB 29 KRGDGSPATWTRGVFTRHSQTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 88
QY 65 GSCLOQRTTTPFLFCNNVNDVCFASNDYSYWLSTPALMPMNPITGRALEPYISRCTV 124
DB 89 GSCLOQRTTTPFLFCNNVNDVCFASNDYSYWLSTPALMPMNPITGRALEPYISRCTV 148
QY 125 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMTSAGSEG 167
DB 149 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMTSAGSEG 191

RESULT 4
US-09-512-563C-46
Sequence 46, Application US/09512563C
Patent No. 6579969
GENERAL INFORMATION:
APPLICANT: Saus, Juan

TITLE OF INVENTION: Goodpasture Binding Protein
FILE REFERENCE: 98-723-A
CURRENT APPLICATION NUMBER: US/09/512,563C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/121,483
PRIOR FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 63
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 46
LENGTH: 211
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match 65.2%; Score 159; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.7e-155;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGVFTRHSQTTAIPSCPECTVPLYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGPATWTRGVFTRHSQTTAIPSCPECTVPLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSCLOQRTTTPFLFCNNVNDVCFASNDYSYWLSTPALMPMNPITGRALEPYIS 120
DB 61 LGTLGSCLOQRTTTPFLFCNNVNDVCFASNDYSYWLSTPALMPMNPITGRALEPYIS 120
QY 121 RCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159
DB 121 RCTVCEGPAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159

RESULT 5
US-08-399-889-25
Sequence 25, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 25
LENGTH: 218
TYPE: PRT
ORGANISM: Human
US-08-399-889-25

Query Match 57.8%; Score 141; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 8.6e-137;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 QTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGLGSCLOQRTTTPFLFCNNVNDV 86
DB 1 QTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGLGSCLOQRTTTPFLFCNNVNDV 60
QY 87 FASNDYSYWLSTPALMPMNPITGRALEPYISRCTVCEGPAIAVHSQTTDIPPCPH 146
DB 61 FASNDYSYWLSTPALMPMNPITGRALEPYISRCTVCEGPAIAVHSQTTDIPPCPH 120

RESULT 6

```
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match          57.8%; Score 141; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 8.6e-137;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFVTMPFLFCNVNDVCN 86
Db 1 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFVTMPFLFCNVNDVCN 60

Qy 87 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
Db 61 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120

Qy 147 GWISLWKGFSPIMFTSAGSEG 167
Db 121 GWISLWKGFSPIMFTSAGSEG 141

RESULT 7
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match          57.8%; Score 141; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 8.6e-137;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFVTMPFLFCNVNDVCN 86
Db 1 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFVTMPFLFCNVNDVCN 60

Qy 87 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
Db 61 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120

Qy 147 GWISLWKGFSPIMFTSAGSEG 167
Db 121 GWISLWKGFSPIMFTSAGSEG 141

US-09-512-563C-50
; Sequence 50, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 50
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-IV-V
US-09-512-563C-50

Query Match          25.0%; Score 61; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 4.1e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60
Db 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60

Qy 61 L 61
Db 61 L 61

RESULT 8
US-09-512-563C-48
; Sequence 48, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 48
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII
US-09-512-563C-48

Query Match          25.0%; Score 61; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 4.4e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60
Db 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60

Qy 61 L 61
Db 61 L 61

RESULT 9
US-09-512-563C-52
; Sequence 52, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 52
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII
US-09-512-563C-52

Query Match          25.0%; Score 61; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 4.4e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60
Db 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60

Qy 61 L 61
Db 61 L 61

RESULT 10
US-09-512-563C-52
; Sequence 52, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 52
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII
US-09-512-563C-52
```

```
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 52
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-V
US-09-512-563C-52

Query Match          25.0%; Score 61; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 4.4e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCDEGTVPLYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCDEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 L 61
DB 61 L 61

RESULT 11
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; PRIOR FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match          16.0%; Score 39; DB 2; Length 471;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 372

RESULT 12
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
```

```
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; PRIOR FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match          16.0%; Score 39; DB 3; Length 471;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 372

RESULT 13
US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match          16.0%; Score 39; DB 3; Length 471;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 372

RESULT 14
US-09-512-563C-61
; Sequence 61, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 61
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-512-563C-61

Query Match          15.2%; Score 37; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.9e-30;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPEG 37
```

Db 1 GLKRGDGGSPATWTRGFVTRHSQTATPSCPEG 37
|||||

RESULT 15

US-09-439-897-65
; Sequence 65, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C8 alpha3
US-09-439-897-65

Query Match 14.8%; Score 36; DB 3; Length 36;

Best Local Similarity 100.0%; Pred. No. 1.1e-29; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 0;

Qy 209 SLNPERMFKPIPTVKAGELEKIISRCQVCKKXH 244
|||||

Db 1 SLNPERMFKPIPTVKAGELEKIISRCQVCKKXH 36
|||||

RESULT 16

US-09-439-897-63
; Sequence 63, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C7 alpha3
US-09-439-897-63

Query Match 10.7%; Score 26; DB 3; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.5e-19; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0;

Qy 1 GLKRGDGGSPATWTRGFVTRHS 26
|||||

Db 1 GLKRGDGGSPATWTRGFVTRHS 26
|||||

RESULT 17

US-09-512-563C-26
; Sequence 26, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C

; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl
US-09-512-563C-26

Query Match 8.6%; Score 21; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.7e-14; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

Qy 3 KGKRGDGGSPATWTRGFVFT 23
|||||

Db 1 KGKRGDGGSPATWTRGFVFT 21
|||||

RESULT 18

US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 7.0%; Score 17; DB 4; Length 260;

Best Local Similarity 100.0%; Pred. No. 2.2e-09; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 84 VCNFASNDYSYWLSTP 100
|||||

Db 101 VCNFASNDYSYWLSTP 117
|||||

RESULT 19

US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2

Query Match 7.0%; Score 17; DB 4; Length 260;

Best Local Similarity 100.0%; Pred. No. 2.2e-09; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

```
QY      84 VCNFASRNDYSYWLSTP 100
Db      101 VCNFASRNDYSYWLSTP 117

RESULT 20
US-09-589-987-2
; Sequence 2, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-2

Query Match      7.0%; Score 17; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      101 VCNFASRNDYSYWLSTP 117

RESULT 21
US-09-589-927-10
; Sequence 10, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-10

Query Match      7.0%; Score 17; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      105 VCNFASRNDYSYWLSTP 121

RESULT 22
US-09-277-665-10
; Sequence 10, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-10

Query Match      7.0%; Score 17; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      105 VCNFASRNDYSYWLSTP 121

RESULT 23
US-09-589-987-10
; Sequence 10, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-10

Query Match      7.0%; Score 17; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      105 VCNFASRNDYSYWLSTP 121

RESULT 24
US-09-439-897-53
; Sequence 53, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C2 alpha3
US-09-439-897-53

Query Match      6.1%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      29 TAIPSCPEGTVPVLYS 43
Db      1 TAIPSCPEGTVPVLYS 15
```


RESULT 25
US-09-439-897-57
; Sequence 57, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
US-09-439-897-57

Query Match 6.1%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EKIISRCQVCKKXKH 244
Db 1 EKIISRCQVCKKXKH 15

RESULT 26
US-09-439-897-59
; Sequence 59, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
US-09-439-897-59

Query Match 6.1%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LPMNMAPITGRALE 116
Db 1 LPMNMAPITGRALE 15

RESULT 27
US-09-439-897-61
; Sequence 61, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 61
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
US-09-439-897-61

Query Match 6.1%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 TDIPPCPHGWISLWK 153
Db 1 TDIPPCPHGWISLWK 15

RESULT 28
US-09-512-563C-27
; Sequence 27, Application US/09512563C
; Patent No. 6579869
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl1a1a9
US-09-512-563C-27

Query Match 5.7%; Score 14; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTTRGVFT 23
Db 8 GSPATWTTRGVFT 21

RESULT 29
US-07-621-091G-3
; Sequence 3, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621,091G

FILING DATE: 11/30/90
CLASSIFICATION: 424
PRIOR APPLICATION DATA: No. 5424408 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Barth, Richard S.
REGISTRATION NUMBER: 28180
REFERENCE/DOCKET NUMBER: 900983/RB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acid residues
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: CDNA to mRNA
DESCRIPTION: Synthetic peptide corresponding to the deduced
DESCRIPTION: amino acid sequence of the carboxy terminal 12 amino acids of SEQ 1
FRAGMENT TYPE: C-terminal fragment
ORIGINAL SOURCE:
ORGANISM: Human
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole human kidney
PUBLICATION INFORMATION: No. 5424408e
US-07-621-091G-3

Query Match 4.9%; Score 12; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKRRH 244
| | | | | | | | | | | | | |
Db 1 ISRCQVCWKRRH 12
RESULT 30
US-08-399-889-3
Sequence 3, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Readers, Stephen T
APPLICANT: Morrison, Karen E
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 3
LENGTH: 12
TYPE: PRT
ORGANISM: Human
US-08-399-889-3

Query Match 4.9%; Score 12; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKRRH 244
| | | | | | | | | | | | | |
Db 1 ISRCQVCWKRRH 12
RESULT 31
US-09-167-364-3
Sequence 3, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:

APPLICANT: Readers, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 3
LENGTH: 12
TYPE: PRT
ORGANISM: Human
US-09-167-364-3

Query Match 4.9%; Score 12; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKRRH 244
| | | | | | | | | | | | | |
Db 1 ISRCQVCWKRRH 12
RESULT 32
US-09-439-897-5
Sequence 5, Application US/09439897
Patent No. 6277558
GENERAL INFORMATION:
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 5
LENGTH: 12
TYPE: PRT
ORGANISM: Homo sapiens
US-09-439-897-5

Query Match 4.9%; Score 12; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCWKRRH 244
| | | | | | | | | | | | | |
Db 1 ISRCQVCWKRRH 12
RESULT 33
US-09-439-897-55
Sequence 55, Application US/09439897
Patent No. 6277558
GENERAL INFORMATION:
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 55
LENGTH: 12
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Chimeric
OTHER INFORMATION: construct C3 alpha
US-09-439-897-55

```
Query Match      4.1%, Score 10; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SLNPERMRKPI 220
| | | | | | | | | |
DB 1 SLNPERMRKPI 12

RESULT 34
US-09-589-927-4
; Sequence 4, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-4

Query Match      4.1%, Score 10; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
| | | | | | | | | |
DB 142 PAIAIAVHSQ 151

RESULT 35
US-09-277-665-4
; Sequence 4, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251-I
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-4

Query Match      4.1%, Score 10; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
| | | | | | | | | |
DB 142 PAIAIAVHSQ 151

RESULT 36
US-09-589-987-4
; Sequence 4, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
```

```
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-4

Query Match      4.1%, Score 10; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137
| | | | | | | | | |
DB 142 PAIAIAVHSQ 151

RESULT 37
US-09-512-563C-44
; Sequence 44, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII-IV-V
US-09-512-563C-44

Query Match      3.3%, Score 8; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61
| | | | | | | |
DB 1 QRAHGQDL 8

RESULT 38
US-09-512-563C-43
; Sequence 43, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII derived
```

; OTHER INFORMATION: peptide

US-09-512-563C-43

Query Match 3.3%; Score 8; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHQDGL 61
| | | | |
Db 1 QRAHQDGL 8

RESULT 39

US-09-589-927-12
; Sequence 12, Application US/09589927
; Patent No. 6432706

; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center

; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251

; CURRENT APPLICATION NUMBER: US/09/589,927

; CURRENT FILING DATE: 2000-06-07

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12

; LENGTH: 260

; TYPE: PRT

; ORGANISM: Human

US-09-589-927-12

Query Match 3.3%; Score 8; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
| | | | |
Db 251 SRCQVCWK 258

RESULT 40

US-09-277-665-12
; Sequence 12, Application US/09277665
; Patent No. 6440729

; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center

; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1

; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12

; LENGTH: 260

; TYPE: PRT

; ORGANISM: Human

US-09-277-665-12

Query Match 3.3%; Score 8; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
| | | | |
Db 251 SRCQVCWK 258

RESULT 41

US-09-589-987-12
; Sequence 12, Application US/09589987
; Patent No. 6498140

; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251

; CURRENT APPLICATION NUMBER: US/09/589,987

; CURRENT FILING DATE: 2000-06-07

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12

; LENGTH: 260

; TYPE: PRT

; ORGANISM: Human

US-09-589-987-12

Query Match 3.3%; Score 8; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
| | | | |
Db 251 SRCQVCWK 258

RESULT 42

US-09-439-897-56

; Sequence 56, Application US/09439897

; Patent No. 6277558

; GENERAL INFORMATION:

; APPLICANT: Hudson, Billy G

; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

; FILE REFERENCE: 95-1263-C

; CURRENT APPLICATION NUMBER: US/09/439,897

; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 56

; LENGTH: 14

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Chimeric

; OTHER INFORMATION: construct C4 alpha1

US-09-439-897-56

Query Match 2.9%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 240
| | | | |
Db 5 SRCQVCWK 11

RESULT 43

US-09-439-897-64

; Sequence 64, Application US/09439897

; Patent No. 6277558

; GENERAL INFORMATION:

; APPLICANT: Hudson, Billy G

; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

; FILE REFERENCE: 95-1263-C

; CURRENT APPLICATION NUMBER: US/09/439,897

; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 64

; LENGTH: 35

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Chimeric

; OTHER INFORMATION: construct C8 alpha1

US-09-439-897-64

```
Query Match      2.9%; Score 7; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCM 240
    |||||
Db 26 SRCQVCM 32

RESULT 44
US-09-134-001C-3093
; Sequence 3093, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 674
; SEQ ID NO 3093
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3093

Query Match      2.9%; Score 7; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSPA 13
    |||||
Db 40 GDSGSPA 46

RESULT 45
US-09-395-689-6
; Sequence 6, Application US/09395689
; Patent No. 6387684
; GENERAL INFORMATION:
; APPLICANT: Hwang, Jaulang
; APPLICANT: Hui, cho-Pat
; APPLICANT: Chen, Tzong-Yueh
; TITLE OF INVENTION: TOPOLISOMERASE 1-MEDIATED DNA DELIVERY
; FILE REFERENCE: 089191/024001
; CURRENT APPLICATION NUMBER: US/09/395,689
; CURRENT FILING DATE: 1999-09-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-395-689-6

Query Match      2.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
    |||||
Db 88 ALASPGS 94

RESULT 46
US-09-411-578-1
; Sequence 1, Application US/09411578
```

```
; Patent No. 6203801
; GENERAL INFORMATION:
; APPLICANT: Schaap, Theodorus C
; APPLICANT: Kuiper, Catharina M
; APPLICANT: Vermeulen, Arnoldus N
; TITLE OF INVENTION: Coccidiosis Vaccines
; FILE REFERENCE: schaap
; CURRENT APPLICATION NUMBER: US/09/411,578
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 98203384.7
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 98203457.1
; EARLIER FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Eimeria tenella
US-09-411-578-1

Query Match      2.9%; Score 7; DB 3; Length 214;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
    |||||
Db 14 ALEPYIS 20

RESULT 47
US-09-749-233-1
; Sequence 1, Application US/09749233
; Patent No. 6680061
; GENERAL INFORMATION:
; APPLICANT: Schaap, Theodorus C
; APPLICANT: Kuiper, Catharina M
; APPLICANT: Vermeulen, Arnoldus N
; TITLE OF INVENTION: Coccidiosis Vaccines
; FILE REFERENCE: schaap
; CURRENT APPLICATION NUMBER: US/09/749,233
; CURRENT FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/411,578
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: 98203457.1
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Eimeria tenella
US-09-749-233-1

Query Match      2.9%; Score 7; DB 4; Length 214;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
    |||||
Db 14 ALEPYIS 20

RESULT 48
US-08-486-099-96
; Sequence 96, Application US/08486099
; Patent No. 6013263
; GENERAL INFORMATION:
; APPLICANT: Bolocnesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
```

```

; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
; TITLE OF INVENTION: B VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-486-099-96

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 49
US-08-360-107A-106
; Sequence 106, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-360-107A-106

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 50
US-08-484-223B-96
; Sequence 96, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE

```

INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-223B-96

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 51

US-08-919-597-96
Sequence 96, Application US/08919597
Patent No. 6054265
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/919,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/470,896
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 869-9741/8864
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-919-597-96

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 52

US-08-475-668A-96
Sequence 96, Application US/08475668A
Patent No. 606065
GENERAL INFORMATION:
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10038-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,668A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 869-9741/8864
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-475-668A-96

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 53

US-08-485-551A-96
Sequence 96, Application US/08485551A
Patent No. 6068973
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE

```

; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,551A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-551A-96

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 54
US-08-471-913A-96
; Sequence 96, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 55
US-08-485-264A-96
; Sequence 96, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids

```



```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-485-264A-96

Query Match      2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      165 SEGTGQA 171
Db      29 SEGTGQA 35

RESULT 56
US-08-474-349A-96
; Sequence 96, Application US/08474349A
; Patent No. 633395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway Jr., Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-474-349A-96

Query Match      2.9%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      165 SEGTGQA 171
Db      29 SEGTGQA 35

RESULT 57
US-08-255-208A-32
; Sequence 32, Application US/08255208A
; Patent No. 643656
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway Jr., Stephen R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-255-208A-32

Query Match      2.9%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      165 SEGTGQA 171
Db      29 SEGTGQA 35

RESULT 58
US-08-470-896-96
; Sequence 96, Application US/08470896
; Patent No. 647955
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
```

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/470,896
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-470-896-96

Query Match 2.9%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 29 SEGTGQA 35

RESULT 59
US-08-485-546A-96
Sequence 96, Application US/08485546A
Patent No. 6518013
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TRANSMISSION
NUMBER OF SEQUENCES: 214
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,546A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-546A-96

Query Match 2.9%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 29 SEGTGQA 35

RESULT 60
PCT-US94-01149-2
Sequence 2, Application PC/TUS9401149
GENERAL INFORMATION:
APPLICANT: Shatzman, Allan
APPLICANT: Scott, Miller
APPLICANT: Dillon, Susan B.
APPLICANT: Kane, James
TITLE OF INVENTION: Vaccinal Polypeptides
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation - Corporate
STREET: U.S. Mailcode UM2220 - 709 Swedeland Road
CITY: King of Prussia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01149
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 149,150
FILING DATE: 05-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 013,415
FILING DATE: 01-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 108,914
FILING DATE: 18-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 837,773
FILING DATE: 18-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 751,896
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 387,200
FILING DATE: 28-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 238,801
FILING DATE: 02-NOV-1988

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 645,732
;; FILING DATE: 30-AUG-1984
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Baumeister, Kirk
;; REGISTRATION NUMBER: 33,833
;; REFERENCE/DOCKET NUMBER: P50134 PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-270-5096
;; TELEFAX: 215-270-5090
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 221 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US94-01149-2

Query Match 2.9%; Score 7; DB 5; Length 221;
Best Local Similarity 100.0%; Pred.No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEGTGQA 171
|||
Db 29 SEGTGQA 35

RESULT 61

PCT-US94-01149-4
;; Sequence 4, Application PC/TUS9401149
;; GENERAL INFORMATION:
;; APPLICANT: Shatzman, Allan
;; APPLICANT: Scott, Miller
;; APPLICANT: Dillon, Susan B.
;; APPLICANT: Kane, James
;; TITLE OF INVENTION: Vaccinal Polypeptides
;; NUMBER OF SEQUENCES: 72
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate
;; ADDRESSEE: Patents
;; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: Pennsylvania
;; COUNTRY: USA
;; ZIP: 19406-2799
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/01149
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 149,150
;; FILING DATE: 05-NOV-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 013,415
;; FILING DATE: 01-FEB-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 108,914
;; FILING DATE: 18-AUG-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 837,773
;; FILING DATE: 18-FEB-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 751,896
;; FILING DATE: 30-AUG-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 387,200
;; FILING DATE: 28-JUL-1989
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 238,801
;; FILING DATE: 02-NOV-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 645,732
;; FILING DATE: 30-AUG-1984
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Baumeister, Kirk
;; REGISTRATION NUMBER: 33,833
;; REFERENCE/DOCKET NUMBER: P50134 PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-270-5096
;; TELEFAX: 215-270-5090
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 221 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US94-01149-4

Query Match 2.9%; Score 7; DB 5; Length 221;
Best Local Similarity 100.0%; Pred.No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEGTGQA 171
|||
Db 29 SEGTGQA 35

RESULT 62

PCT-US94-01149-59
;; Sequence 59, Application PC/TUS9401149
;; GENERAL INFORMATION:
;; APPLICANT: Shatzman, Allan
;; APPLICANT: Scott, Miller
;; APPLICANT: Dillon, Susan B.
;; APPLICANT: Kane, James
;; TITLE OF INVENTION: Vaccinal Polypeptides
;; NUMBER OF SEQUENCES: 72
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate
;; ADDRESSEE: Patents
;; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: Pennsylvania
;; COUNTRY: USA
;; ZIP: 19406-2799
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/01149
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 149,150
;; FILING DATE: 05-NOV-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 013,415
;; FILING DATE: 01-FEB-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 108,914
;; FILING DATE: 18-AUG-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 837,773
;; FILING DATE: 18-FEB-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 751,896
;; FILING DATE: 30-AUG-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 387,200

```

; FILING DATE: 28-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 238,801
; FILING DATE: 02-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 238,801
; FILING DATE: 02-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 645,732
; FILING DATE: 30-AUG-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50134 PCT
; TELEPHONE: 215-270-5096
; TELEFAX: 215-270-5090
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01149-59

```

```

Query Match          2.9%; Score 7; DB 5; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

```

RESULT 63

```

PCT-US94-01149-60
; Sequence 60, Application PC/TUS9401149
; GENERAL INFORMATION:
; APPLICANT: Shatzman, Allan
; APPLICANT: Scott, Miller
; APPLICANT: Dillon, Susan B.
; APPLICANT: Kane, James
; TITLE OF INVENTION: Vaccinal Polypeptides
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation - Corporate
; ADDRESSEE: Patents
; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 149,150
; FILING DATE: 05-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 013,415
; FILING DATE: 01-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 108,914
; FILING DATE: 18-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 837,773
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 751,896
; FILING DATE: 30-AUG-1991

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 387,200
; FILING DATE: 28-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 238,801
; FILING DATE: 02-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 645,732
; FILING DATE: 30-AUG-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50134 PCT
; TELEPHONE: 215-270-5096
; TELEFAX: 215-270-5090
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01149-60

```

```

Query Match          2.9%; Score 7; DB 5; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

```

RESULT 64

```

PCT-US94-01149-8
; Sequence 8, Application PC/TUS9401149
; GENERAL INFORMATION:
; APPLICANT: Shatzman, Allan
; APPLICANT: Scott, Miller
; APPLICANT: Dillon, Susan B.
; APPLICANT: Kane, James
; TITLE OF INVENTION: Vaccinal Polypeptides
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation - Corporate
; ADDRESSEE: Patents
; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 149,150
; FILING DATE: 05-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 013,415
; FILING DATE: 01-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 108,914
; FILING DATE: 18-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 837,773
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 751,896
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 387,200
FILING DATE: 28-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 238,801
FILING DATE: 02-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 645,732
FILING DATE: 30-AUG-1984
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50134 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-270-5096
TELEFAX: 215-270-5090
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
PCT-US94-01149-8

Query Match 2.9%; Score 7; DB 5; Length 222;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 65
US-08-928-692-31
Sequence 31, Application US/08928692
Patent No. 5958727
GENERAL INFORMATION:
APPLICANT: Brody, Howard
APPLICANT: Yaver, Deborah S.
APPLICANT: Lamsa, Michael
APPLICANT: Hansen, Kim
TITLE OF INVENTION: Methods for Modifying the Production of
TITLE OF INVENTION: a Polypeptide
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,692
FILING DATE: 12-SEPT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
PCT-US94-01149-8

LENGTH: 233 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5958727e
US-08-928-692-31

Query Match 2.9%; Score 7; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 39 ALEPYIS 45

RESULT 66
US-09-339-972-31
Sequence 31, Application US/09339972
Patent No. 6323002
GENERAL INFORMATION:
APPLICANT: Brody, Howard
APPLICANT: Yaver, Deborah S.
APPLICANT: Lamsa, Michael
APPLICANT: Hansen, Kim
TITLE OF INVENTION: Methods for Modifying the Production of
TITLE OF INVENTION: a Polypeptide
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6323002o No. 6323002disk of No. 6323002th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/339,972
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,692
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6323002e
US-09-339-972-31

Query Match 2.9%; Score 7; DB 4; Length 233;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 39 ALEPYIS 45

RESULT 67

```
5223425-2
; Patent No. 5223425
; APPLICANT: FLIER, JEFFREY S.; SPIEGELMAN, BRUCE M.; ROSEN,
; BARRY M.; WHITE, TYLER R.
; TITLE OF INVENTION: DNA ENCODING HUMAN ADIPSIN WITH COMPLEMENT
; D ACTIVITY
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION NUMBER: US/07/277,963
; FILING DATE: 30-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 34,203
; FILING DATE: 02-APR-1987
; SEQ ID NO:2
; LENGTH: 259
5223425-2
Query Match      2.9%; Score 7; DB 6; Length 259;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      6 RGDGSGP 12
      |||||
Db      206 RGDGSGP 212
      |||||

RESULT 68
US-09-589-927-8
; Sequence 8, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-8
Query Match      2.9%; Score 7; DB 6; Length 259;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      6 RGDGSGP 12
      |||||
Db      206 RGDGSGP 212
      |||||

5223425-10
Query Match      2.9%; Score 7; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      174 SPGSCLE 180
      |||||
Db      188 SPGSCLE 194
      |||||

RESULT 71
5223425-10
; Patent No. 5223425
; APPLICANT: FLIER, JEFFREY S.; SPIEGELMAN, BRUCE M.; ROSEN,
; BARRY M.; WHITE, TYLER R.
; TITLE OF INVENTION: DNA ENCODING HUMAN ADIPSIN WITH COMPLEMENT
; D ACTIVITY
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/277,963
; FILING DATE: 30-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 34,203
; FILING DATE: 02-APR-1987
; SEQ ID NO:10
; LENGTH: 260
5223425-10
Query Match      2.9%; Score 7; DB 6; Length 260;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      6 RGDGSGP 12
      |||||
Db      207 RGDGSGP 213
      |||||

RESULT 72
PCT-US94-01149-10
; Sequence 10, Application PC/TUS9401149
; GENERAL INFORMATION:
; APPLICANT: Shatzman, Allan
; APPLICANT: Scott, Miller
; APPLICANT: Dillon, Susan B.
; APPLICANT: Kane, James
; TITLE OF INVENTION: Vaccinal Polypeptides
```

NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation - Corporate
ADDRESSEE: Patents
STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
CITY: King of Prussia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01149
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 149,150
FILING DATE: 05-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 013,415
FILING DATE: 01-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 108,914
FILING DATE: 18-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 837,773
FILING DATE: 18-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 751,896
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 387,200
FILING DATE: 26-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 238,801
FILING DATE: 02-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 645,732
FILING DATE: 30-AUG-1984
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50134 PCT
TELEPHONE: 215-270-5096
TELEFAX: 215-270-5090
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01149-10

Query Match 2.9%; Score 7; DB 5; Length 306;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 114 SEGTGQA 120

RESULT 73
US-08-840-713-4
Sequence 4, Application US/08840713
Patent No. 6498233
GENERAL INFORMATION:
APPLICANT: WELLS, Winfried, Dr.
ATTORNEY/AGENT INFORMATION:
NAME: FOYMINAYA, Jesus

TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
STREET: 655 15th St., N.W., Suite 330 - G St. Lobby
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,713
FILING DATE: 25-APR-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kitts, Monica Chin
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: 1614-7014
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-840-713-4

Query Match 2.9%; Score 7; DB 4; Length 342;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177
Db 136 ALASPGS 142

RESULT 74
US-08-229-781-58
Sequence 58, Application US/08229781
Patent No. 5589174
GENERAL INFORMATION:
APPLICANT: Yoshinobu OKUNO et al.
TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/229,781
FILING DATE: April 19, 1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/054,016
FILING DATE: April 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:
LENGTH: 347 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:

ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:

LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:

NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:

AUTHORS:

TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

Query Match 2.9%; Score 7; DB 1; Length 347;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
Db 155 SEGTGQA 161

RESULT 75

US-08-630-918-58
Sequence 58, Application US/08630918
Patent No. 5631350
GENERAL INFORMATION:

APPLICANT: Yoshinobu OKINO et al.

TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,918
FILING DATE: April 5, 1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,781
FILING DATE: April 19, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/054,016
FILING DATE: April 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 347 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-630-918-58

Query Match 2.9%; Score 7; DB 1; Length 347;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||
Db 155 SEGTGQA 161

RESULT 76

US-09-004-422-58

Sequence 58, Application US/09004422

Patent No. 6337070

GENERAL INFORMATION:

APPLICANT: Yoshinobu OKINO et al.

TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., #800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/004,422

FILING DATE: January 8, 1998

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/443,862

FILING DATE: May 22, 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,781

FILING DATE: April 19, 1994

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/054,016
; FILING DATE: April 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-09-004-422-58

Query Match 2.9%; Score 7; DB 4; Length 347;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 155 SEGTGQA 161

RESULT 77
US-09-973-963-4
; Sequence 4, Application US/09973963
; Patent No. 6653102
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.

APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/973,963
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: US 60/240,790
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: US 60/304,775
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-973-963-4
Query Match 2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 SGSPATW 15
Db 55 SGSPATW 61
RESULT 78
US-09-046-992-4
; Sequence 4, Application US/09046992
; Patent No. 6140066
; GENERAL INFORMATION:
; APPLICANT: Lorboum-Galski, Haya
; APPLICANT: Yarkoni, Shai
; APPLICANT: Ben-Yehudah, Ahmi
; TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS
; TITLE OF INVENTION: USING A CHIMERIC TOXIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,992
; FILING DATE: 24-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 9457-0013-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal

US-09-046-992-4

Query Match 2.9%; Score 7; DB 3; Length 396;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||
Db 98 ALASPGS 104

RESULT 79

US-08-391-259-2

; Sequence 2, Application US/08391259

; Patent No. 5621078

; GENERAL INFORMATION:

; APPLICANT: Riemen, Mark W

; APPLICANT: Stirdivant, Steven M

; TITLE OF INVENTION: Modified PE40

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merck & Co., Inc.

; STREET: 126 Lincoln Avenue

; CITY: Rahway

; STATE: New Jersey

; COUNTRY: U.S.

; ZIP: 07065

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/391,259

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/120,698

; FILING DATE:

; APPLICATION NUMBER: US/07/879,037

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Grassler, Frank P

; REGISTRATION NUMBER: 31,164

; REFERENCE/DOCKET NUMBER: 178791A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (908) 594-3462

; TELEFAX: (908) 594-4720

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 420 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-391-259-2

Query Match

2.9%; Score 7; DB 1; Length 420;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||
Db 122 ALASPGS 128

RESULT 80

US-08-391-259-7

; Sequence 7, Application US/08391259

; Patent No. 5621078

; GENERAL INFORMATION:

; APPLICANT: Riemen, Mark W

; APPLICANT: Stirdivant, Steven M

; TITLE OF INVENTION: Modified PE40

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merck & Co., Inc.

; STREET: 126 Lincoln Avenue

; CITY: Rahway

; STATE: New Jersey

; COUNTRY: U.S.

; ZIP: 07065

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/391,259

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/120,698

; FILING DATE:

; APPLICATION NUMBER: US/07/879,037

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Grassler, Frank P

; REGISTRATION NUMBER: 31,164

; REFERENCE/DOCKET NUMBER: 178791A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (908) 594-3462

; TELEFAX: (908) 594-4720

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 420 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-391-259-7

Query Match

2.9%; Score 7; DB 1; Length 420;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||
Db 122 ALASPGS 128

RESULT 81

US-08-391-259-10

; Sequence 10, Application US/08391259

; Patent No. 5621078

; GENERAL INFORMATION:

; APPLICANT: Riemen, Mark W

; APPLICANT: Stirdivant, Steven M

; TITLE OF INVENTION: Modified PE40

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merck & Co., Inc.

; STREET: 126 Lincoln Avenue

; CITY: Rahway

; STATE: New Jersey

; COUNTRY: U.S.

; ZIP: 07065

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/391,259

; FILING DATE:

; CLASSIFICATION: 530

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,698
; FILING DATE:
; APPLICATION NUMBER: US/07/879,037
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Grassler, Frank P
; REGISTRATION NUMBER: 31,164
; REFERENCE/DOCKET NUMBER: 178791A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3462
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-391-259-10

Query Match          2.9%; Score 7; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      171 ALASPGS 177
Db      122 ALASPGS 128

RESULT 82
US-08-391-259-11
; Sequence 11, Application US/08391259
; Patent No. 5621078
; GENERAL INFORMATION:
; APPLICANT: Riemen, Mark W
; APPLICANT: Stirdivant, Steven M
; TITLE OF INVENTION: Modified PE40
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/391,259
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/120,698
; FILING DATE:
; APPLICATION NUMBER: US/07/879,037
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Grassler, Frank P
; REGISTRATION NUMBER: 31,164
; REFERENCE/DOCKET NUMBER: 178791A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3462
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
US-08-391-259-11

Query Match          2.9%; Score 7; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      171 ALASPGS 177
Db      122 ALASPGS 128

RESULT 83
US-08-839-425-2
; Sequence 2, Application US/08839425
; Patent No. 5912322
; GENERAL INFORMATION:
; APPLICANT: Riemen, Mark W
; APPLICANT: Stirdivant, Steven M
; TITLE OF INVENTION: Modified PE40
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Power Macintosh 6.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,425
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Grassler, Frank P
; REGISTRATION NUMBER: 31,164
; REFERENCE/DOCKET NUMBER: 178791A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3462
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-839-425-2

Query Match          2.9%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      171 ALASPGS 177
Db      122 ALASPGS 128

RESULT 84
US-08-839-425-7
; Sequence 7, Application US/08839425
; Patent No. 5912322
; GENERAL INFORMATION:
; APPLICANT: Riemen, Mark W
; APPLICANT: Stirdivant, Steven M
; TITLE OF INVENTION: Modified PE40
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
```

STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: U.S.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Power Macintosh 6.0.1
SOFTWARE: Microsoftword 6.0.1
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Grassler, Frank P
REGISTRATION NUMBER: 31,164
REFERENCE/DOCKET NUMBER: 178791A
TELEPHONE: (908) 594-3462
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-425-7

Query Match 2.9%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
DB 122 ALASPGS 128

RESULT 85
US-08-839-425-10
Sequence 10, Application US/08839425
Patent No. 5912322
GENERAL INFORMATION:
APPLICANT: Riemen, Mark W
APPLICANT: Stirdivant, Steven M
TITLE OF INVENTION: Modified PE40
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: U.S.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Power Macintosh 6.0.1
SOFTWARE: Microsoftword 6.0.1
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Grassler, Frank P
REGISTRATION NUMBER: 31,164
REFERENCE/DOCKET NUMBER: 178791A
TELEPHONE: (908) 594-3462
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:

LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-425-10

Query Match 2.9%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
DB 122 ALASPGS 128

RESULT 86
US-08-839-425-11
Sequence 11, Application US/08839425
Patent No. 5912322
GENERAL INFORMATION:
APPLICANT: Riemen, Mark W
APPLICANT: Stirdivant, Steven M
TITLE OF INVENTION: Modified PE40
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: U.S.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Power Macintosh 6.0.1
SOFTWARE: Microsoftword 6.0.1
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Grassler, Frank P
REGISTRATION NUMBER: 31,164
REFERENCE/DOCKET NUMBER: 178791A
TELEPHONE: (908) 594-3462
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-425-11

Query Match 2.9%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
DB 122 ALASPGS 128

RESULT 87
US-08-840-713-6
Sequence 6, Application US/08840713
Patent No. 6498233
GENERAL INFORMATION:
APPLICANT: WELS, Winfried, Dr.
APPLICANT: FOYMINAYA, Jesus

;; TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM
;; NUMBER OF SEQUENCES: 58
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nikolaïdo, Marmelstein, Murray & Oram LLP
;; STREET: 655 15th St., N.W., Suite 330 - G St. Lobby
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-5701
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/840,713
;; FILING DATE: 25-APR-1997
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kitts, Monica Chin
;; REGISTRATION NUMBER: 36,105
;; REFERENCE/DOCKET NUMBER: 1614-7014
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 638 - 5000
;; TELEFAX: (202) 638 - 4810
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 421 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-840-713-6

Query Match 2.9%; Score 7; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 215 ALASPGS 221

RESULT 88
US-08-484-438-42
;; Sequence 42, Application US/08484438
;; Patent No. 5811098
;; Patent No. 5811098 5780031
;; GENERAL INFORMATION:
;; APPLICANT: Plowman, Gregory D.
;; APPLICANT: Culouscou, Jean-Michel
;; APPLICANT: Shoyab, Mohammed
;; APPLICANT: Siegall, Clay B.
;; APPLICANT: Heilstr m, Ingegerd
;; APPLICANT: Heilstr m, Karl E.
;; TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
;; NUMBER OF SEQUENCES: 42
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036-2711
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/484,438
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/323,442
;; FILING DATE: 14-OCT-1994
;; APPLICATION NUMBER: US 08/150,704
;; FILING DATE: 10-NOV-1993
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/981,165
;; FILING DATE: 24-NOV-1992
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Misrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 5624-230
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864/9741
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 42:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 462 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-484-438-42

Query Match 2.9%; Score 7; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 164 ALASPGS 170

RESULT 89
US-08-840-713-2
;; Sequence 2, Application US/08840713
;; Patent No. 6498233
;; GENERAL INFORMATION:
;; APPLICANT: WELS, Winfried, Dr.
;; APPLICANT: FOYMINAYA, Jesus
;; TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM
;; NUMBER OF SEQUENCES: 58
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nikolaïdo, Marmelstein, Murray & Oram LLP
;; STREET: 655 15th St., N.W., Suite 330 - G St. Lobby
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-5701
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/840,713
;; FILING DATE: 25-APR-1997
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kitts, Monica Chin
;; REGISTRATION NUMBER: 36,105
;; REFERENCE/DOCKET NUMBER: 1614-7014
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 638 - 5000
;; TELEFAX: (202) 638 - 4810
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 530 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-840-713-2

Query Match 2.9%; Score 7; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
|||||||
DB 324 ALASPGS 330

RESULT 90
US-09-506-286B-8
; Sequence 8, Application US/09506286B
; Patent No. 6482414
; GENERAL INFORMATION:
; APPLICANT: Dowling, Patricia W.
; APPLICANT: Youngner, Julius S.
; APPLICANT: The University of Pittsburgh, of the Commonwealth
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES
; FILE REFERENCE: EQ-1-C2
; CURRENT APPLICATION NUMBER: US/09/506,286B
; CURRENT FILING DATE: 2000-02-16
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: PCT/US99/18583
; PRIOR FILING DATE: 1999-08-12
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Equine influenza virus H3N8
US-09-506-286B-8

Query Match 2.9%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||||
DB 373 SEGTGQA 379

RESULT 91
US-09-506-286B-11
; Sequence 11, Application US/09506286B
; Patent No. 6482414
; GENERAL INFORMATION:
; APPLICANT: Dowling, Patricia W.
; APPLICANT: Youngner, Julius S.
; APPLICANT: The University of Pittsburgh, of the Commonwealth
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES
; FILE REFERENCE: EQ-1-C2
; CURRENT APPLICATION NUMBER: US/09/506,286B
; CURRENT FILING DATE: 2000-02-16
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: PCT/US99/18583
; PRIOR FILING DATE: 1999-08-12
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Equine influenza virus H3N8
US-09-506-286B-11

Query Match 2.9%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||||

DB 373 SEGTGQA 379

RESULT 92
US-09-762-861B-8
; Sequence 8, Application US/09762861B
; Patent No. 6579528
; GENERAL INFORMATION:
; APPLICANT: The University of Pittsburgh - of the Commonwealth System of Higher
; APPLICANT: Education
; APPLICANT: Dowling, Patricia W.
; APPLICANT: Youngner, Julius S.
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES
; FILE REFERENCE: EQ-1-C1-PUS (formerly HKZ-033CPUS)
; CURRENT APPLICATION NUMBER: US/09/762,861B
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: PCT/US99/18583
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 09/133,921
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Equine influenza virus H3N8
US-09-762-861B-8

Query Match 2.9%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||||
DB 373 SEGTGQA 379

RESULT 93
US-09-762-861B-11
; Sequence 11, Application US/09762861B
; Patent No. 6579528
; GENERAL INFORMATION:
; APPLICANT: The University of Pittsburgh - of the Commonwealth System of Higher
; APPLICANT: Education
; APPLICANT: Dowling, Patricia W.
; APPLICANT: Youngner, Julius S.
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES
; FILE REFERENCE: EQ-1-C1-PUS (formerly HKZ-033CPUS)
; CURRENT APPLICATION NUMBER: US/09/762,861B
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: PCT/US99/18583
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 09/133,921
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Equine influenza virus H3N8
US-09-762-861B-11

Query Match 2.9%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
|||||||
DB 373 SEGTGQA 379

RESULT 94
US-10-065-133A-8

; Sequence 8, Application US/10065133A
; Patent No. 6685946
; GENERAL INFORMATION:
; APPLICANT: Dowling, Patricia W.
; APPLICANT: Youngner, Julius S.
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES
; FILE REFERENCE: EQ-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/065,133A
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US99/18583
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 09/133,921
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Equine influenza virus H3N8
US-10-065-133A-8

Query Match 2.9%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 95
US-10-065-133A-11
; Sequence 11, Application US/10065133A
; Patent No. 6685946
; GENERAL INFORMATION:
; APPLICANT: Dowling, Patricia W.
; APPLICANT: Youngner, Julius S.
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES
; FILE REFERENCE: EQ-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/065,133A
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US99/18583
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 09/133,921
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Equine influenza virus H3N8
US-10-065-133A-11

Query Match 2.9%; Score 7; DB 4; Length 565;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 96
US-09-232-468A-22
; Sequence 22, Application US/09232468A
; Patent No. 6207165
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/232,468A
; CURRENT FILING DATE: 1999-01-05

; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 22
; LENGTH: 566
; TYPE: PRT
; ORGANISM: swine influenza virus
US-09-232-468A-22

Query Match 2.9%; Score 7; DB 3; Length 566;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 374 SEGTGQA 380

RESULT 97
US-09-784-984B-53
; Sequence 53, Application US/09784984B
; Patent No. 6576243
; GENERAL INFORMATION:
; APPLICANT: Meriel Ltd.
; APPLICANT: Audonnet, Jean-Christophe
; APPLICANT: Bouchardon, Annabelle
; APPLICANT: Baudu, Philippe
; APPLICANT: Riviere, Michael
; TITLE OF INVENTION: Polynucleotide Vaccine Formula Against Porcine Reproductive and
; FILE REFERENCE: 454313-2230.1
; CURRENT APPLICATION NUMBER: US/09/784,984B
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: FR 96/09338
; PRIOR FILING DATE: 1996-07-19
; PRIOR APPLICATION NUMBER: PCT/FR97/01313
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: US 6,207,165
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 53
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Swine Influenza Virus
US-09-784-984B-53

Query Match 2.9%; Score 7; DB 4; Length 566;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 374 SEGTGQA 380

RESULT 98
US-08-453-848-7
; Sequence 7, Application US/08453848
; Patent No. 5858368
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; FILE REFERENCE: HEMAGGLUTININ MULTIVALENT VACCINES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta

```
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/453,848
  FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/120,607
  FILING DATE: 13-SEPT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: Pabst, Patrea L.
  REGISTRATION NUMBER: 31,284
  REFERENCE/DOCKET NUMBER: MGS101CIP
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (404)-873-8794
    TELEFAX: (404)-873-8795
  INFORMATION FOR SEQ ID NO: 7:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 570 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: Peptide
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE: N-terminal
      ORIGINAL SOURCE:
        ORGANISM: Influenza virus
        INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
      FEATURE:
        NAME/KEY: AcNPV 61K protein signal sequence
        LOCATION: 1 to 18
      FEATURE:
        NAME/KEY: mature rHA
        LOCATION: 19 to 552
US-08-453-848-7

Query Match 2.9%; Score 7; DB 2; Length 570;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 378 SEGTGQA 384

RESULT 99
US-09-169-027-7
Sequence 7, Application US/09169027
Patent No. 6245532
GENERAL INFORMATION:
  APPLICANT: Smith, Gale Eugene
  APPLICANT: Volvovitz, Franklin
  APPLICANT: Wilkinson, Bethanie Eident
  APPLICANT: Voznesensky, Andrei I.
  APPLICANT: Hackett, Craig Stanway
  TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
  TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Patrea L. Pabst
    STREET: 2800 One Atlantic Center
    STREET: 1201 West Peachtree Street
    CITY: Atlanta
    STATE: GA
    COUNTRY: USA
    ZIP: 30309-3450

STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/453,848
  FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/120,607
  FILING DATE: 13-SEPT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: Pabst, Patrea L.
  REGISTRATION NUMBER: 31,284
  REFERENCE/DOCKET NUMBER: MGS101CIP
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (404)-873-8794
    TELEFAX: (404)-873-8795
  INFORMATION FOR SEQ ID NO: 7:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 570 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: Peptide
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE: N-terminal
      ORIGINAL SOURCE:
        ORGANISM: Influenza virus
        INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
      FEATURE:
        NAME/KEY: AcNPV 61K protein signal sequence
        LOCATION: 1 to 18
      FEATURE:
        NAME/KEY: mature rHA
        LOCATION: 19 to 552
US-08-453-848-7

Query Match 2.9%; Score 7; DB 2; Length 570;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 378 SEGTGQA 384

RESULT 100
US-08-453-848-15
Sequence 15, Application US/08453848
Patent No. 5858368
GENERAL INFORMATION:
  APPLICANT: Smith, Gale Eugene
  APPLICANT: Volvovitz, Franklin
  APPLICANT: Wilkinson, Bethanie Eident
  APPLICANT: Voznesensky, Andrei I.
  APPLICANT: Hackett, Craig Stanway
  TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
  TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Patrea L. Pabst
    STREET: 2800 One Atlantic Center
    STREET: 1201 West Peachtree Street
    CITY: Atlanta
    STATE: GA
    COUNTRY: USA
    ZIP: 30309-3450
```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,848
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/120,607
; FILING DATE: 13-SEPT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MGS101CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-873-8794
; TELEFAX: (404)-873-8795
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORGANISM: Influenza virus
; INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
; FEATURE:
; NAME/KEY: AcNPV 61K protein signal sequence
; LOCATION: 1 to 18
; NAME/KEY: mature rHA
; LOCATION: 19 to 553
; US-08-453-848-15

```

```

Query Match      2.9%; Score 7; DB 2; Length 571;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      165 SEGTGQA 171
Db      379 SEGTGQA 385

RESULT 101
US-08-453-848-21
; Sequence 21, Application US/08453848
; Patent No. 5858368
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,848
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/120,607
; FILING DATE: 13-SEPT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MGS101CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-873-8794
; TELEFAX: (404)-873-8795
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORGANISM: Influenza virus
; INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA
; FEATURE:
; NAME/KEY: AcNPV 61K protein signal sequence
; LOCATION: 1 to 18
; NAME/KEY: mature rHA
; LOCATION: 19 to 569
; US-08-453-848-21

```

```

Query Match      2.9%; Score 7; DB 2; Length 571;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      165 SEGTGQA 171
Db      379 SEGTGQA 385

RESULT 102
US-09-169-027-15
; Sequence 15, Application US/09169027
; Patent No. 6245532
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/169,027
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/453,848
;; FILING DATE: 30-MAY-1995
;; APPLICATION NUMBER: 08/120,607
;; FILING DATE: 13-SEPT-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Pabst, Patrea L.
;; REGISTRATION NUMBER: 31,284
;; REFERENCE/DOCKET NUMBER: MGS101CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (404)-873-8794
;; TELEFAX: (404)-873-8795
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 571 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;; ORIGINAL SOURCE:
;; ORGANISM: Influenza virus
;; INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
;; FEATURE:
;; NAME/KEY: AcNPV 61K protein signal sequence
;; LOCATION: 1 to 18
;; FEATURE:
;; NAME/KEY: mature rHA
;; LOCATION: 19 to 553
;; US-09-169-027-15

Query Match 2.9%; Score 7; DB 3; Length 571;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTOQA 171
Db 379 SEGTOQA 385

RESULT 103
US-09-169-027-21
; Sequence 21, Application US/09169027
; Patent No. 6245532
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volnovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/169,027
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/453,848
;; FILING DATE: 30-MAY-1995
;; APPLICATION NUMBER: 08/120,607
;; FILING DATE: 13-SEPT-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Pabst, Patrea L.
;; REGISTRATION NUMBER: 31,284
;; REFERENCE/DOCKET NUMBER: MGS101CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (404)-873-8794
;; TELEFAX: (404)-873-8795
;; INFORMATION FOR SEQ ID NO: 21:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 571 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;; ORIGINAL SOURCE:
;; ORGANISM: Influenza virus
;; INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA
;; FEATURE:
;; NAME/KEY: AcNPV 61K protein signal sequence
;; LOCATION: 1 to 18
;; FEATURE:
;; NAME/KEY: mature rHA
;; LOCATION: 19 to 569
;; US-09-169-027-21

Query Match 2.9%; Score 7; DB 3; Length 571;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTOQA 171
Db 379 SEGTOQA 385

RESULT 104
US-08-463-163-3
; Sequence 3, Application US/08463163
; Patent No. 5696237
; GENERAL INFORMATION:
; APPLICANT: Fitzgerald, David J.
; APPLICANT: Chaudhary, Vijay K.
; APPLICANT: Pastan, Ira H.
; APPLICANT: Waldmann, Thomas A.
; APPLICANT: Queen, Cary L.
; TITLE OF INVENTION: Recombinant Antibody-Toxin Fusion Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,163
; FILING DATE: 05-JUN-1995

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/227,227
; FILING DATE: 22-JAN-1981
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/911,227
; FILING DATE: 24-SEP-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,361
; FILING DATE: 21-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865,722
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-12211
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-463-163-3

Query Match 2.9%; Score 7; DB 1; Length 599;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 301 ALASPGS 307

RESULT 105
US-08-405-615-1
; Sequence 1, Application US/08405615
; Patent No. 5602095
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Fitzgerald, David J.
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with
; TITLE OF INVENTION: Increased Activity
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ellen L. Weber
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,615
; FILING DATE:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,709
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 613 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-615-1

Query Match 2.9%; Score 7; DB 1; Length 613;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 315 ALASPGS 321

RESULT 106
US-08-461-234-1
; Sequence 1, Application US/08461234
; Patent No. 5821238
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira H.
; APPLICANT: Fitzgerald, David J.
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with
; TITLE OF INVENTION: Increased Activity
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,234
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,709
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,615
; FILING DATE: 15-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-36-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 613 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; APPLICATION TYPE: NO
; HYPOTHETICAL: NO
; US-08-461-234-1

Query Match 2.9%; Score 7; DB 2; Length 613;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/297,851
FILING DATE: 30-JUL-1999
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030,376
FILING DATE: 06-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/20207
FILING DATE: 05-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Storella, John R.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 015280-29810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-297-851-2

Query Match 2.9%; Score 7; DB 4; Length 613;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177
Db 315 ALASPGS 321

RESULT 110
US-08-225-224-1
Sequence 1, Application US/08225224
Patent No. 5635599
GENERAL INFORMATION:
APPLICANT: PASTAN, Ira
APPLICANT: KREITMAN, Robert J.
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,224
FILING DATE: 8-APR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen L.
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 614 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein

FEATURE:
NAME/KEY: Protein
LOCATION: 1..614
OTHER INFORMATION: /label= native-PE
US-08-225-224-1
Query Match 2.9%; Score 7; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 171 ALASPGS 177
Db 316 ALASPGS 322
RESULT 111
US-08-722-258-1
Sequence 1, Application US/08722258
Patent No. 6011002
GENERAL INFORMATION:
APPLICANT: Pastan, Ira
APPLICANT: Kreicman, Robert J.
APPLICANT: Puri, Raj K.
TITLE OF INVENTION: Circularly Permuted Ligands and
TITLE OF INVENTION: Circularly Permuted Chimeric Molecules
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,258
FILING DATE: 08-JAN-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/04468
FILING DATE: 06-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/225,224
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 015280-193100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 614 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..614
OTHER INFORMATION: /note= "native Pseudomonas exotoxin
OTHER INFORMATION:
US-08-722-258-1
Query Match 2.9%; Score 7; DB 3; Length 614;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 316 ALASPGS 322

RESULT 112
PCT-US95-04468-1
; Sequence 1, Application PC/TUS9504468
; GENERAL INFORMATION:
; APPLICANT: CIRCULARLY PERMUTATED LIGANDS AND
; TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
; NUMBER OF SEQUENCES: 59
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04468
; FILING DATE: 07-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,224
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-193-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 614 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..614
; OTHER INFORMATION: /label= native-PE
PCT-US95-04468-1

Query Match 2.9%; Score 7; DB 5; Length 614;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 316 ALASPGS 322

RESULT 113
US-09-970-516-4
; Sequence 4, Application US/09970516
; Patent No. 6610534
; GENERAL INFORMATION:
; APPLICANT: NO. 6610534artis AG
; TITLE OF INVENTION: Induction of blood vessel formation through administration of
; FILE REFERENCE: 4-31617
; CURRENT APPLICATION NUMBER: US/09/970,516
; CURRENT FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-970-516-4

Query Match 2.9%; Score 7; DB 4; Length 618;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 395 ALASPGS 401

RESULT 114
US-08-356-786-16
; Sequence 16, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppermann, Hermann
; APPLICANT: Houston, L. L.
; APPLICANT: Ring, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; TITLE OF INVENTION: Marker
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; STREET: Exchange Place, 53 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,786
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/831,967
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-356-786-16

Query Match 2.9%; Score 7; DB 2; Length 622;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 324 ALASPGS 330

RESULT 115
US-09-046-992-2
; Sequence 2, Application US/09046992
; Patent No. 6140066
; GENERAL INFORMATION:
; APPLICANT: Lorberboun-Galski, Haya
; APPLICANT: Yarkoni, Shai
; APPLICANT: Ben-Yehudah, Ahmi
; TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS

;; TITLE OF INVENTION: USING A CHIMERIC TOXIN
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds, LLP
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10036-2811
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; OPERATING SYSTEM: Windows
;; SOFTWARE: FastSeq for Windows Version 2.0b
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/046,992
;; FILING DATE: 24-MAR-1998
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Poissant, Brian M
;; REGISTRATION NUMBER: 28,462
;; REFERENCE/DOCKET NUMBER: 9457-0013-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-493-4935
;; TELEFAX: 650-493-5556
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 635 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: Single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FRAGMENT TYPE: internal
;; US-09-046-992-2

Query Match 2.9%; Score 7; DB 3; Length 635;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177
|||
Db 337 ALASPGS 343

RESULT 116
US-08-235-838-14
; Sequence 14, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/235,838
;; FILING DATE: TBA
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/828,832
;; FILING DATE: 31-JAN-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: GB 91-810079.3
;; FILING DATE: 05-FEB-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919)541-8614
;; TELEFAX: (919)541-8689
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 637 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-235-838-14

Query Match 2.9%; Score 7; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177
|||
Db 339 ALASPGS 345

RESULT 117
US-08-235-838-16
; Sequence 16, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,838
; FILING DATE: TBA
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 637 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-235-838-16

Query Match 2.9%; Score 7; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 92; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

Qy 171 ALASPGS 177
Db 339 ALASPGS 345

RESULT 118
US-08-465-473B-14
; Sequence 14, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hessa J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-473B-14

Query Match 2.9%; Score 7; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 92; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

Qy 171 ALASPGS 177
Db 339 ALASPGS 345

RESULT 119
US-08-465-473B-16
; Sequence 16, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hessa J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-473B-16

Query Match 2.9%; Score 7; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 92; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

Qy 171 ALASPGS 177
Db 339 ALASPGS 345

RESULT 120
US-09-047-148-2
; Sequence 2, Application US/09047148
; Patent No. 6086900

| | | | | | | | | | |
|--|-----|--------------|-----|------------|----|--------|----|------|----|
| Matches | 7; | Conservative | 0; | Mismatches | 0; | Indels | 0; | Gaps | 0; |
| QY | 174 | SPGSCL | 180 | | | | | | |
| | | | | | | | | | |
| DB | 429 | SPGSCL | 435 | | | | | | |
| RESULT 122 | | | | | | | | | |
| US-08-398-590A-40 | | | | | | | | | |
| ; Sequence 40, Application US/08398590A | | | | | | | | | |
| ; Patent No. 5935811 | | | | | | | | | |
| ; GENERAL INFORMATION: | | | | | | | | | |
| ; APPLICANT: Anderson, David J. | | | | | | | | | |
| ; APPLICANT: Schoenherr, Christopher J. | | | | | | | | | |
| ; TITLE OF INVENTION: Neuron-Restrictive Silencer Factor | | | | | | | | | |
| ; TITLE OF INVENTION: Proteins | | | | | | | | | |
| ; NUMBER OF SEQUENCES: 53 | | | | | | | | | |
| ; CORRESPONDENCE ADDRESS: | | | | | | | | | |
| ; ADDRESS: Flehr, Hobbach, Test, Albritton & Harbert | | | | | | | | | |
| ; STREET: Four Embarcadero Center, Suite 3400 | | | | | | | | | |
| ; CITY: San Francisco | | | | | | | | | |
| ; STATE: California | | | | | | | | | |
| ; COUNTRY: United States | | | | | | | | | |
| ; ZIP: 94111-4187 | | | | | | | | | |
| ; COMPUTER READABLE FORM: | | | | | | | | | |
| ; MEDIUM TYPE: Floppy disk | | | | | | | | | |
| ; COMPUTER: IBM PC compatible | | | | | | | | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | | | | | | | | |
| ; SOFTWARE: Patent in Release #1.0, Version #1.30 | | | | | | | | | |
| ; CURRENT APPLICATION DATA: | | | | | | | | | |
| ; APPLICATION NUMBER: US/08/398,590A | | | | | | | | | |
| ; FILING DATE: 03-MAR-1995 | | | | | | | | | |
| ; CLASSIFICATION: 435 | | | | | | | | | |
| ; PRIOR APPLICATION DATA: | | | | | | | | | |
| ; APPLICATION NUMBER: US 08/103,445 | | | | | | | | | |
| ; FILING DATE: 06-AUG-1995 | | | | | | | | | |
| ; ATTORNEY/AGENT INFORMATION: | | | | | | | | | |
| ; NAME: Silva, Robin M. | | | | | | | | | |
| ; REGISTRATION NUMBER: 38,304 | | | | | | | | | |
| ; REFERENCE/DOCKET NUMBER: A-60897/RFT/RMS | | | | | | | | | |
| ; TELECOMMUNICATION INFORMATION: | | | | | | | | | |
| ; TELEPHONE: (415) 781-1989 | | | | | | | | | |
| ; TELEFAX: (415) 398-3249 | | | | | | | | | |
| ; TELEX: 910 277299 | | | | | | | | | |
| ; INFORMATION FOR SEQ ID NO: 40: | | | | | | | | | |
| ; SEQUENCE CHARACTERISTICS: | | | | | | | | | |
| ; LENGTH: 676 amino acids | | | | | | | | | |
| ; TYPE: amino acid | | | | | | | | | |
| ; TOPOLOGY: linear | | | | | | | | | |
| ; MOLECULE TYPE: protein | | | | | | | | | |
| US-08-398-590A-40 | | | | | | | | | |
| Query Match 2.9%; Score 7; DB 2; Length 676; | | | | | | | | | |
| Best Local Similarity 100.0%; Pred. No. 97; | | | | | | | | | |
| Matches | 7; | Conservative | 0; | Mismatches | 0; | Indels | 0; | Gaps | 0; |
| QY | 7 | GDSGSPA | 13 | | | | | | |
| | | | | | | | | | |
| DB | 64 | GDSGSPA | 70 | | | | | | |
| RESULT 123 | | | | | | | | | |
| US-08-894-997-40 | | | | | | | | | |
| ; Sequence 40, Application US/08894997A | | | | | | | | | |
| ; Patent No. 6270990 | | | | | | | | | |
| ; GENERAL INFORMATION: | | | | | | | | | |
| ; APPLICANT: Anderson, David J. | | | | | | | | | |
| ; APPLICANT: Schoenherr, Christopher J. | | | | | | | | | |
| ; TITLE OF INVENTION: NEURON-RESTRICTIVE SILENCER FACTOR | | | | | | | | | |
| ; FILE REFERENCE: 17810-502 NRSF | | | | | | | | | |
| ; CURRENT APPLICATION NUMBER: US/08/894,997A | | | | | | | | | |
| ; CURRENT FILING DATE: 1998-01-06 | | | | | | | | | |
| ; EARLIER APPLICATION NUMBER: PCT/US96/02817 | | | | | | | | | |

; EARLIER FILING DATE: 1996-03-01
; EARLIER APPLICATION NUMBER: 08/398,590
; EARLIER FILING DATE: 1995-03-03
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CHAIN
; LOCATION: (1)..(676)
; OTHER INFORMATION: Human NSRF (partial)
US-08-894-997-40

Query Match 2.9%; Score 7; DB 3; Length 676;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDSGSPA 13
Db 64 GDSGSPA 70

RESULT 124

US-09-252-991A-19361
; Sequence 19361, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19361
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19361

Query Match 2.9%; Score 7; DB 4; Length 767;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177
Db 469 ALASPGS 475

RESULT 125

US-08-494-168-2
; Sequence 2, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1694 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-494-168-2

Query Match 2.9%; Score 7; DB 1; Length 1694;
Best Local Similarity 100.0%; Pred. No. 2.3e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 SRCQVCV 240
Db 1676 SRCQVCV 1682

RESULT 126

US-09-679-279-14
; Sequence 14, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Volchegursky, Yanina
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; TITLE OF INVENTION: Genes and Uses Thereof
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 3562
; TYPE: PRT
; ORGANISM: Micromonospora megalomicea
US-09-679-279-14

Query Match 2.9%; Score 7; DB 4; Length 3562;
Best Local Similarity 100.0%; Pred. No. 4.5e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 GTVPLYS 43
Db 2228 GTVPLYS 2234

RESULT 127

US-09-439-897-51
; Sequence 51, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:

APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 51
LENGTH: 6
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Chimeric
OTHER INFORMATION: construct C1 alpha3
US-09-439-897-51

Query Match 2.5%; Score 6; DB 3; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATWTR 18
Db 1 ATWTR 6

RESULT 128
US-10-080-505-58
Sequence 58, Application US/10080505
Patent No. 6676948
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RFT/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/080,505
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 58
LENGTH: 7
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-080-505-58

Query Match 2.5%; Score 6; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
Db 1 GDSGSP 6

RESULT 129
US-08-296-791-7
Sequence 7, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION/DOCKET NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-296-791-7

Query Match 2.5%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
Db 1 GDSGSP 6

RESULT 130
US-08-296-791-8
Sequence 8, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION/DOCKET NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-296-791-8

Query Match 2.5%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
|||||
Db 1 GDSGSP 6

RESULT 131

US-09-839-996-7
; Sequence 7, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/839,996

FILING DATE: 20-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791

FILING DATE: 25-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-839-996-7

Query Match 2.5%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
|||||
Db 1 GDSGSP 6

RESULT 132

US-09-839-996-8
; Sequence 8, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/839,996

FILING DATE: 20-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791

FILING DATE: 25-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-839-996-8

Query Match 2.5%; Score 6; DB 4; Length 8;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12

Db 1 GDSGSP 6

RESULT 133

US-10-080-505-53
; Sequence 53, Application US/10080505
; Patent No. 6678948
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-53

Query Match 2.5%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
|||||
Db 1 GDSGSP 6

Db 1 GDSGSP 6

RESULT 134

US-10-080-505-54
; Sequence 54, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DC7/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 54
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-080-505-54

Query Match 2.5%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 7 GDSGSP 12

Db 1 GDSGSP 6

RESULT 135

PCT-US95-10661A-7
; Sequence 7, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
PCT-US95-10661A-7

Query Match 2.5%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 7 GDSGSP 12

Db 1 GDSGSP 6

RESULT 136

PCT-US95-10661A-8
; Sequence 8, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
PCT-US95-10661A-8

Query Match 2.5%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 7 GDSGSP 12

Db 1 GDSGSP 6

RESULT 137

US-09-322-624-4
; Sequence 4, Application US/09322624
; Patent No. 6548734
; GENERAL INFORMATION:
; APPLICANT: Glimcher, L et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO MODULATION OF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; CURRENT APPLICATION NUMBER: US/09/322,624

; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: USSN 09/087,139
; EARLIER FILING DATE: 1998-03-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 4
; LENGTH: 9
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-322-624-4

Query Match 2.5%; Score 6; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIA 133
Db 2 PAIAIA 7

RESULT 138
US-10-080-505-20
; Sequence 20, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:

; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-20

Query Match 2.5%; Score 6; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDSGSP 12
Db 1 GDSGSP 6

RESULT 139
US-08-260-582-40
; Sequence 40, Application US/08260582
; Patent No. 5635182
; GENERAL INFORMATION:

; APPLICANT: McCoy, John M.
; APPLICANT: Lu, Zhijian
; TITLE OF INVENTION: METHOD OF DETECTING LIGAND
; TITLE OF INVENTION: INTERACTIONS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/260,582
; FILING DATE: 16-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, M. C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5236
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-260-582-40

Query Match 2.5%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TQQAIA 173
Db 3 TQQAIA 8

RESULT 140

PCT-US95-05471-40
; Sequence 40, Application PC/TUS9505471
; GENERAL INFORMATION:

; APPLICANT:
; TITLE OF INVENTION: METHOD OF DETECTING LIGAND INTERACTIONS
; NUMBER OF SEQUENCES: 76
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05471
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-05471-40

Query Match 2.5%; Score 6; DB 5; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TQQAIA 173
Db 3 TQQAIA 8

RESULT 141

US-09-439-897-58
; Sequence 58, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C

```
/ CURRENT APPLICATION NUMBER: US/09/439,897
/ CURRENT FILING DATE: 1999-11-12
/ NUMBER OF SEQ ID NOS: 65
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 58
/ LENGTH: 15
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Chimeric
/ OTHER INFORMATION: construct C5 alpha
US-09-439-897-58

Query Match          2.5%; Score 6; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITG 112
Db      |||||
        6 MAPITG 11

RESULT 142
US-09-106-568E-149
/ Sequence 149, Application US/09106568E
/ Patent No. 6455248
/ GENERAL INFORMATION:
/ APPLICANT: Bhattacherjee, J.
/ APPLICANT: Suvarna, Kalavati
/ APPLICANT: Bhattacherjee, Vasker
/ TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
/ TITLE OF INVENTION: A BIOLOGICAL SAMPLE
/ FILE REFERENCE: 96,247-A
/ CURRENT APPLICATION NUMBER: US/09/106,568E
/ CURRENT FILING DATE: 1998-05-29
/ PRIOR APPLICATION NUMBER: 08/650,809
/ PRIOR FILING DATE: 1997-05-20
/ NUMBER OF SEQ ID NOS: 160
/ SOFTWARE: Microsoft Word 97
/ SEQ ID NO 149
/ LENGTH: 16
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Polypeptide segment of LYS2_CALB shown in Figure 4.
US-09-106-568E-149

Query Match          2.5%; Score 6; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 TTAIPS 33
Db      |||||
        10 TTAIPS 15

RESULT 143
US-08-716-249-1
/ Sequence 1, Application US/08716249
/ Patent No. 6455244
/ GENERAL INFORMATION:
/ APPLICANT: Guichard, Gilles
/ APPLICANT: Muller, Sylviane
/ APPLICANT: Briand, Jean-Paul
/ APPLICANT: Regemortel, Marc
/ TITLE OF INVENTION: Retropeptides, Antibodies Thereto, and
/ TITLE OF INVENTION: Uses Thereof for Vaccination and In Vitro Diagnosis
/ NUMBER OF SEQUENCES: 13
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Spencer & Frank
/ STREET: 1100 New York Avenue, Suite 300E
/ CITY: Washington, D.C.
/ COUNTRY: USA
```

```
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/716,249
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/00292
FILING DATE: 13-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Calvetti, Frederick F.
REGISTRATION NUMBER: 28,557
REFERENCE/DOCKET NUMBER: GROFO 7001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)414-4000
TELEFAX: (202)414-4040
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-716-249-1

Query Match          2.5%; Score 6; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSGS 11
Db      |||||
        6 RGDGSGS 11

RESULT 144
US-08-716-249-9
/ Sequence 9, Application US/08716249
/ Patent No. 6455244
/ GENERAL INFORMATION:
/ APPLICANT: Guichard, Gilles
/ APPLICANT: Muller, Sylviane
/ APPLICANT: Briand, Jean-Paul
/ APPLICANT: Regemortel, Marc
/ TITLE OF INVENTION: Retropeptides, Antibodies Thereto, and
/ TITLE OF INVENTION: Uses Thereof for Vaccination and In Vitro Diagnosis
/ NUMBER OF SEQUENCES: 13
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Spencer & Frank
/ STREET: 1100 New York Avenue, Suite 300E
/ CITY: Washington, D.C.
/ COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/716,249
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/00292
FILING DATE: 13-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Calvetti, Frederick F.
REGISTRATION NUMBER: 28,557
REFERENCE/DOCKET NUMBER: GROFO 7001
```

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)414-4000
TELEFAX: (202)414-4040
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: YES
ANTI-SENSE: YES
US-08-716-249-9

Query Match 2.5%; Score 6; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSGS 11

Db 6 RGDGSGS 11

RESULT 145

US-08-449-287-20
Sequence 20, Application US/08449287
Patent No. 5877293

GENERAL INFORMATION:

APPLICANT: ADAIR, John Robert
APPLICANT: BODMER, Mark William
APPLICANT: MOUNTAIN, Andrew
APPLICANT: OWENS, Raymond John
TITLE OF INVENTION: CDR Grafted Anti-CEA Antibodies and
TITLE OF INVENTION: Their Production
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,287
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/154,389

FILING DATE:

APPLICATION DATA:

APPLICATION NUMBER: PCT GB91/01108

FILING DATE: 05-JUL-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9014932.9

FILING DATE: 05-JUL-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT GB90/02017

FILING DATE: 21-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: SAXE, Bernhard D.

REGISTRATION NUMBER: 28,665

REFERENCE/DOCKET NUMBER: 40283/110 CARA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-449-287-20

Query Match

2.5%; Score 6; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134

Db 5 AIAIAV 10

RESULT 146

US-09-003-081-8
Sequence 8, Application US/09003081
Patent No. 5988779

GENERAL INFORMATION:

APPLICANT: Campfield, Arthur Dr.
APPLICANT: Devos, Rene Dr.
APPLICANT: Guisez, Yves Dr.
TITLE OF INVENTION: Recombinant Obese (OB) Proteins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche, Inc.
STREET: 340 Kingeland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/003,081
FILING DATE:
CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/435,777

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Picut, Catherine A

REGISTRATION NUMBER: 37419

REFERENCE/DOCKET NUMBER: 9165

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-4387

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-09-003-081-8

Query Match

2.5%; Score 6; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134

Db 5 AIAIAV 10

RESULT 147

US-08-648-262-8
Sequence 8, Application US/08648262
Patent No. 6025324

GENERAL INFORMATION:

APPLICANT: Bailon, Pascal Mr.
APPLICANT: Campfield, Arthur Dr.
APPLICANT: Devos, Rene Dr.
APPLICANT: Guisez, Yves Dr.
TITLE OF INVENTION: Pegylated Obese (OB) Proteins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche, Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/648,262
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Epstein, William H.
REGISTRATION NUMBER: 20008
REFERENCE/DOCKET NUMBER: 9281
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-3723
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-648-262-8

Query Match 2.5%; Score 6; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
Db 5 AIAIAV 10

RESULT 148
US-08-648-263-8
Sequence 8, Application US/08/648263
Patent No. 6025325
GENERAL INFORMATION:
APPLICANT: Campfield, Arthur
APPLICANT: Devos, Rene
APPLICANT: Guisez, Yves
TITLE OF INVENTION: RECOMBINANT OBSE (OB) PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/648,263

FILING DATE: 15-MAY-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/484,629
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/435,777
FILING DATE: 05-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kreisler, Lewis J.
REGISTRATION NUMBER: 38522
REFERENCE/DOCKET NUMBER: RAN 4105/175-002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4387
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-648-263-8

Query Match 2.5%; Score 6; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
Db 5 AIAIAV 10

RESULT 149
US-08-840-713-48
Sequence 48, Application US/08840713
Patent No. 6498233
GENERAL INFORMATION:
APPLICANT: WELS, Winfried, Dr.
APPLICANT: FOYMINAVA, Jesus
TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
STREET: 655 15th St., N.W., Suite 330 - G St. Lobby
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,713
FILING DATE: 25-APR-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kitts, Monica Chin
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: 1614-7014
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638 - 5000
TELEFAX: (202) 638 - 4810
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-840-713-48

Query Match 2.5%; Score 6; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
Db 5 AIAIAV 10

RESULT 150

US-09-904-196B-3
; Sequence 3, Application US/09904196B
; Patent No. 6555660
; GENERAL INFORMATION:
; APPLICANT: NISSEN, TORBEN LAURSGAARD
; APPLICANT: ANDERSEN, KIM VILBOUR
; APPLICANT: HANSEN, CHRISTIAN KARSTEN
; APPLICANT: MIKKELSEN, JAN MOLLER
; TITLE OF INVENTION: G-CSF CONJUGATES
; FILE REFERENCE: 31-000700US
; CURRENT APPLICATION NUMBER: US/09/904,196B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US/09/760,008
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/176,376
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/189,506
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 60/215,644
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DK PA 2000 00024
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: DK PA 2000 00341
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: DK PA 2000 00943
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-904-196B-3

Query Match 2.5%; Score 6; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
Db 5 AIAIAV 10

Search completed: April 5, 2004, 07:39:59
Job time : 26 secs